# BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number
WO 01/09352 A2

(51) International Patent Classification[7]: C12N 15/80, 15/10, 15/67, 1/14 // (C12N 1/14, C12R 1:66)

(21) International Application Number: PCT/NL00/00544

(22) International Filing Date: 28 July 2000 (28.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99202516.3 30 July 1999 (30.07.1999) EP
99202517.1 30 July 1999 (30.07.1999) EP

(71) Applicants *(for all designated States except US)*: **NEDERLANDSE ORGANISATIE VOOR TOEGEPAST-NATUURWETENSCHAPPELIJK ONDERZOEK (TNO)** [NL/NL]; Schoemakerstraat 97, NL-2628 VK Delft (NL). **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HEERIKHUISEN, Margreet** [NL/NL]; Johan van Oldenbarneveltlaan 52-1, NL-3705 HH Zeist (NL). **VAN DEN HONDEL, Cornelis** [NL/NL]; Waterlelie 124, NL-2804 PZ Gouda (NL). **PUNT, Peter** [NL/NL]; Boekdrukkersgilde 1, NL-3994 XT Houten (NL). **VAN BIEZEN, Nick** [NL/NL]; Futenweide 16, NL-3993 DR Houten (NL). **ALBERS, Alwin** [NL/NL]; Laan van Vollenhove 1535, NL-3706 GB Zeist (NL). **VOGEL, Kurt** [CH/CH]; 41 Eulerstrasse, CH-4051 Basel (CH).

(74) Agent: **JORRITSMA, Ruurd**; Nederlandsch Octrooibureau, Scheveningseweg 82, P.O. Box 29720, NL-2502 LS The Hague (NL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *Without international search report and to be republished upon receipt of that report.*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/09352 A2

(54) Title: NOVEL MEANS OF TRANSFORMATION OF FUNGI AND THEIR USE FOR HETEROLOGOUS PROTEIN PRODUCTION

(57) Abstract: A recombinant *Aspergillus sojae* comprising an introduced acetamidase S (*amdS*) gene as selectable marker is disclosed. An *Aspergillus sojae* exhibiting growth with medium comprising uracil and fluoro-orotic acid, said *Aspergillus sojae* further not exhibiting growth on medium comprising uridine and fluoro-orotic acid i.e. said *Aspergillus sojae* exhibiting uracil auxotrophy, said *Aspergillus sojae* being unable to utilize uridine, said *Aspergillus sojae* being pyrG negative, said *Aspergillus sojae* exhibiting resistance to fluoro-orotic acid, said uracil auxotrophy and said fluoro-orotic acid resistance being relievable upon complementation with an active introduced *pyrG* gene, is described. *The Aspergillus sojae* further comprises a nucleic acid sequence encoding a phytase or a protein having phytase activity or any other heterologous protein or polypeptide and can be used for the biotechnological production of said phytase or said other heterologous proteins or polypeptides. Additional mutants exhibiting amended morphology are also disclosed. Methods of producing such expression hosts are described.

# Novel means of transformation of fungi and their use for heterologous protein production.

## SUMMARY OF THE INVENTION

The invention relates to novel means of transformation of fungi and to their use for production of heterologous proteins. The means involve genetically engineered fungi belonging to the taxonomic group *Aspergillus sojae*. Suggestions have been provided in the past to use *Aspergillus sojae* as a host strain for transformation. However to date no data are provided on successful transformation and/or expression of heterologous proteins. In addition it has been found, that so far certain proteins, such as phytase which were difficult to express in large amounts, due to several reasons including proteolytic degradation in expression hosts other than *Aspergillus sojae,* can surprisingly be expressed in *Aspergillus sojae*. Production levels for heterologous proteins in *Aspergillus sojae* have been found to exceed those levels achieved for the same proteins in *Aspergillus niger* and *Aspergillus awamori*. In addition to the above, the subject of the invention further covers a process for obtaining improved *Aspergillus sojae* strains for expression purposes, characterized by, on the one hand, a decreased proteolytic activity, and, on the other hand improved fermentation characteristics related to the morphology of the fungi.

## BACKGROUND OF THE INVENTION

Suggestions have been provided in the past to use *Aspergillus sojae* as a host for transformation. However, to date no data are provided on successful transformation and/or production of heterologous proteins and, more specifically, nothing is revealed concerning expression of phytase. Previously, expression levels were too low in expression hosts other than *Aspergillus sojae*, mainly due to proteolytic degradation. We have now found expression levels for the protein in *Aspergillus sojae* that exceed those levels achieved for the same protein in other strains, e.g. *Aspergillus niger*, *Aspergillus awamori* and *Trichoderma*. It is surprising to find such an improvement in closely related strains. Thus, prior art disclosures concerning phytase production exhibit shortcomings. Prior art disclosures on the use of *Aspergillus sojae* for expressing heterologous proteins or polypeptides were inadequate.

The fact that until now hardly any successful attempts for *A. sojae* transformation have been reported is remarkable in view of the fact that numerous successful transformations of closely related strains of the taxonomic group *Aspergillus oryzae* have

been described in the past. On the basis of this close relationship the skilled person would anticipate (and in fact did anticipate) that analogous methods to those used for *Aspergillus oryzae* are also applicable for *Aspergillus sojae* strains.

WO97/04108 for example describes the isolation of a protease encoding nucleic acid sequence, specifically a leucine aminopeptidase encoding sequence, and the transformation of a variety of host organisms, i.a. *Aspergillus sojae*, with a leucine aminopeptidase encoding sequence. However no illustration of this particular transformation actually having been carried out is provided. It is merely suggested as one possibility among many other strains such as *Trichoderma reesei*, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus phoenicis*, and *Aspergillus oryzae* as a potential host strain to be used for transformation. In the cited document, 3 transformation protocols readily used in the art are suggested for the strains. Specifically any of the selection markers acetamidase S (=amdS) (e.g. as maintained on vector p3SR2), argB or hygromycin B (e.g. using the vector pAN7-1) are suggested as being suitable markers to be used according to the transformation protocols described therein.

The use of vector p3SR2 with the amdS marker has frequently been described in the literature as being useful for transforming various strains, for example *Aspergillus oryzae* (in EP 0.238.023), *Trichoderma reesei* (in EP 0.244.234) and *Aspergillus niger* (EMBO Journal 4, pages 475-479). Consequently, the analogous use for transforming *Aspergillus* in general is put forward in WO97/04108 on the basis of these previous publications.

Quite specifically on page 17 of WO97/04108 it is described that *Aspergilli* and *Trichoderma* that prior to transformation grew slowly on minimal medium comprising solely the substrate acetamide as source of nitrogen could be selected after transformation with the vector p3SR2 due to a clear growth advantage. Subsequently, the thus obtained transformants would need to be further subjected to selection for leucine aminopeptidase (=LAP) productivity in order to find a desired transformant. As stated above this is merely put forward as speculative means of transformation applicable over the two aforementioned genera in toto based on a few successful transformations of strains other than *Aspergillus sojae*.

The suggested transformation protocol is, however, unsuccessful with *Aspergillus sojae*. The selection criteria described in the prior art are insufficient to ensure practical selection of desirable transformants when using the vector p3SR2. We have conducted the

experiments and found the described method inoperable due to excessive background growth eliminating practical selectability.

Another routinely used selection method for fungal transformants is that of transformation of orotidine-5-monophosphate decarboxylase (=PyrG) mutants. Mattern *et al.* in Mol. Gen. Genet. 210, pages 460-461 disclose transformation of *Aspergillus oryzae* using the *Aspergillus niger pyrG* gene. Standard practice is to isolate pyrG mutants based on direct resistance to fluoro-orotic acid as a positive selection marker. This has resulted in isolation of numerous pyrG mutants for a variety of fungi to date.

From experience with a number of different filamentous fungi, the auxotrophic pyrG-based system has many favourable characteristics. Experiments were carried out to obtain *A. sojae* pyrG mutant strains, using a standard procedure based on direct selection for resistance to fluoro-orotic acid (FOA) on plates containing uridine to support growth of the mutant strain (Van Hartingsveldt *et al.* in Mol. Gen. Genet. (1987) 206, pages 71-75). However, use of the analogous method on *Aspergillus sojae* strains did not lead to pyrG mutants. The usual method did lead to fluoro-orotic acid resistant strains but all the strains were able to grow without uridine. Thus, none of these strains were pyrG mutants. Normally, the isolation of the pyrG mutants can be done directly from the fluoro-orotic acid resistant strains on a uridine selection medium. For *Aspergillus sojae* this method however turned out to be inoperable.

Clearly, *Aspergillus sojae* exhibits different traits than the closely related *Aspergillus oryzae* when it comes down to transformation. The standard protocols using amdS or pyrG as selectable markers do not suffice. Unfortunately, the method of argB as selectable marker is not an attractive option either, since this requires isolation of a corresponding argB mutant for every host strain one wishes to use. This is an arduous task based on trial and error. The required argB mutant can be obtained through random mutagenesis followed by screening of tens of thousands of colonies. The situation for pyrG is better in that the mutant is itself selectable. In the case of amdS no mutant is required as the presence of amdS works as dominant selectable marker.

Additional problems dissuading the skilled person from use of *Aspergillus sojae* as expression host for recombinant proteins or polypeptides exist. In JP-A-02-234666 for example an ArgB based selection of *Aspergillus sojae* is described using an analogous protocol to that described for other fungi. Such a process has been described for *Aspergillus oryzae* in Biotechnology (1988) 6, pages 1419-1422. The cited article also refers to successful analogous transformation of *Aspergillus nidulans* and *Aspergillus*

*niger*. However, when the *Aspergillus sojae* strain ATCC42251 disclosed in the Japanese patent application was analysed, an undesirable protease profile was found. The protease profile of this strain is incompatible for application as a production host. So even though a transformation protocol has been suggested in the prior art for this particular *Aspergillus sojae* strain it could not possibly lead to a high level of expression of heterologous protein even if the protocol for transformation was successful.

It is in fact due precisely to the explicit characteristics of *Aspergillus sojae* strains to produce excessive amounts of alkaline proteases and amylases that they currently find application in practice. They are used specifically in processes requiring degradation of complex polymeric substrates. It was thus at best to be expected that any transformants of *Aspergillus sojae* that are finally successful will not lead to good expression levels unless the product is an *Aspergillus sojae* protein that is impervious to its own proteases.

In summary the problems facing the skilled person in finding a means to use *Aspergillus sojae* strains for expressing heterologous recombinant proteins on an industrial scale are manifold. Firstly, a number of processes for introducing the desired nucleic acid material to be expressed are not applicable in the manner used for other fungi. This includes pyrG- and amdS-based processes that are useful for the closely related *Aspergillus oryzae*. Secondly, it remained to be seen whether high level production of heterologous proteins would be feasible despite the known excessive proteolytic activity of the host strain *Aspergillus sojae*.

Unexpectedly, it has been found that the problems addressed above can be solved, thus resulting in novel expression hosts for producing proteins and novel methods of production of heterologous proteins. We describe transformation of *A. sojae* strains with the amdS and pyrG selection markers. In addition efficient gene expression is described, including expression of a phytase gene.

## DESCRIPTION OF THE INVENTION

As stated the subject invention is directed at *Aspergillus sojae* strains and the application thereof for production of recombinant proteins and polypeptides. Firstly, a description of *Aspergillus sojae* strains is provided.

### *Aspergillus sojae* determination.

The fungal taxonomy is a complex issue. The *Aspergillus* genus comprises *Aspergillus sojae* in the Flavi/Tamarii section (see Table 1). *A. sojae* is clearly shown to be

distinct from *A. oryzae* which is located in the same section (see Table 2). Currently, strains belonging to *Aspergillus sojae* can be distinguished from taxonomically closely related *Aspergillus oryzae* and also closely related *Aspergillus parasiticus* strains in a number of manners recognised in the art. Reference is made to the random PCR fragments, *ver-1*, *aflR* and rDNA sequences as described, respectively, in Ushijama *et al.* (1981), Chang *et al.* (1995), Yuan *et al.* (1995), Kusomoto *et al.* (1998) and Watson *et al.* (1999). In addition it has been found that *Aspergillus oryzae* further differs from *Aspergillus sojae* upon comparison of the *alpA* sequence of these strains. Inter alia (there are other sequence differences between *A. oryzae* and *A. sojae alpA* which could be used as a determination tool), it has been found that *Aspergillus sojae* comprises an *Xmn*I restriction site at a specific location in the *alpA* gene. The corresponding location in the *alpA* gene of several *Aspergillus oryzae* strains does not possess such a restriction site. Thus, this provides an additional discrimination point between the two types of fungal strains. Consequently, numerous methods are available to the skilled person to assess whether a strain is an *Aspergillus sojae*. Currently more than 10 strains are deposited with the ATCC that are defined as *Aspergillus sojae*. The 10 oldest deposits have been analysed. Two out of 10 did not pass the lastly mentioned determination test. One of them is the ATCC20235 which according to Ushijama *et al.* (1981) also did not fulfil the requirements for classification as an *Aspergillus sojae* on the basis of morphological parameters. The other is ATCC46250. The definition of *Aspergillus sojae* as used throughout the patent application is meant to imply a strain that preferably fulfils all the requirements described in the cited references in combination with the presence of the *Xmn*I restriction site in the *alpA* gene. Specific homologous primers for both the *Aspergillus oryzae* and *Aspergillus sojae* sequences are also provided. They can be used to test for the presence of the *Xmn*I restriction site by way of example of a screening test useful for distinguishing *Aspergillus oryzae* from *Aspergillus sojae* (Primer sequences are SEQ ID No.1 MBL1784: 5'-CGGAATTCGAGCGCAACTACAAGATCAA-3' and SEQ ID No.2 MBL1785: 5'-CGGAATTCAGCCCAGTTGAAGCCGTC-3'). They are derived from the coding region of the *alpA* gene. It will be obvious to the skilled person on the basis of the known sequence data that alternative probes or primers are conceivable. PCR amplification using these primers on *Aspergillus* DNA, followed by restriction enzyme digestion of the resulting DNA fragments with *Xmn*I provides a way to discriminate *A. sojae* strains from *A. oryzae* strains. Having established the definition of *Aspergillus sojae* strains we can proceed further with the detailed description of the invention.

The invention in one aspect covers a recombinant *Aspergillus sojae* comprising an introduced *acetamidase S (amdS)* gene as a selectable marker. Such an *A. sojae* is selectable on a medium comprising a substrate for the introduced amdS protein as sole source of nitrogen, said medium further comprising a carbon substrate and said medium being free of endogenous amdS inducing substrate. A suitable medium comprises acrylamide as substrate for the introduced amdS as sole source of nitrogen. A suitable medium at least further comprises minimum substrates required for growth of *Aspergillus sojae*. A suitable category of *A. sojae* according to the invention is formed by *A. sojae* that are not selectable on acetamide comprising medium. An *A. sojae* according to the invention is suitably an *A. sojae* selectable on a medium free of glucose, i.e. a medium wherein the carbon source is not glucose. Such a medium can be a medium having sorbitol as carbon source. Best results in the case of sorbitol are achieved when sorbitol is the sole carbon source.

An *Aspergillus sojae* according to the invention may comprise a further introduced nucleic acid sequence, said further introduced sequence preferably encoding a protein or polypeptide. The further introduced sequence may be adapted for optimised codon usage to the host strain codon usage or may have the original codons from the host from which it has been derived. The introduced sequence is in principle any sequence the skilled person wishes to express. The introduced sequence can suitably be heterologous, i.e. foreign to the *Aspergillus sojae* into which it is introduced. It can also be native but introduced in the form of one or more additional copies.

One of the subjects of the invention is aimed at expressing phytase or proteins having phytase activity. Numerous sequences are known to the skilled person concerning sequence data of phytases. We refer to and incorporate by reference the contents of EP 684.313, EP 897.010, WO 99/49022, EP 911.416 and EP 897.985. These documents describe various natural and modified phytase sequences. They also describe a consensus sequence. A suitable embodiment is formed by phytase sequences from *Peniophora* being either the natural sequences or modified versions thereof. The new system is more flexible than prior systems and thus heterologous sequences, including heterologous sequences encoding phytase or proteins having phytase activity that were difficult to express in the prior art fungal systems can be expressed in the novel system according to the invention.

An *Aspergillus sojae* according to the invention as defined in any of the embodiments defined above comprising an introduced *amdS* gene as selectable marker may suitably have no active endogenous *amdS* gene. The *Aspergillus sojae* according to such an embodiment

may by way of example have an endogenous *amdS* gene comprising an endogenous amdS inactivating mutation. Any type of inactivating mutation known or conceivable to the skilled person may have occurred. A suitable example of such inactivating mutation may be a deletion or disruption. The mutation may inactivate the gene or the gene product. The skilled person will realise that numerous options are available to achieve this and that they can readily be achieved.

In an alternative embodiment the invention is also directed at a recombinant *Aspergillus sojae* free of an active endogenous *amdS* gene and further comprising an introduced *amdS* gene as selectable marker. The recombinant *Aspergillus sojae* according to the invention is selectable on a medium comprising a substrate for the amdS as sole source of nitrogen, said medium further comprising a carbon substrate. A suitable medium at least further comprises minimum substrates required for growth of *A. sojae*. In a suitable embodiment the endogenous *amdS* gene can for example have been inactivated. This inactivation can be any type of inactivation known or conceivable to a person skilled in the art that still leaves the *A. sojae* viable. By way of example the endogenous *amdS* gene can comprise an inactivating mutation in the form of a substitution, deletion or insertion of the gene or part thereof, or by virtue of a mutation affecting expression of the gene such as to render it inactive. The complete endogenous *amdS* gene can also be absent.

An *Aspergillus sojae* in any of the described embodiments according to the invention may be an *A. sojae* into an *amdS* gene has been introduced. This can be achieved e.g. by transformation or transfection. The resulting *Aspergillus sojae* according to the invention must then subsequently have been separated from non transformed or transfected *A. sojae*. Any of the embodiments described above as such or in combination are covered by the invention.

The invention not only covers *Aspergillus sojae* as such, but also covers a method of introducing a nucleic acid sequence into *A. sojae*. The method comprises subjecting *Aspergillus sojae* to introduction of a nucleic acid sequence in a manner known per se for introduction of a nucleic acid sequence into a fungus. Such a manner can e.g. be transformation or transfection of the *A. sojae*. The method comprises the introduction of the *amdS* gene as the nucleic acid sequence followed by selection of the resulting transformed or transfected *A. sojae* on a medium free of endogenous amdS inducing substrate, said medium further comprising a substrate for the introduced amdS as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired *A. sojae* comprising introduced *amdS* gene to grow whilst eliminating growth of *A. sojae*

devoid of a functional *amdS* gene. A suitable embodiment of such a method involves applying a medium comprising a substrate for amdS other than acetamide. Suitably, such a medium comprises acrylamide as substrate for the introduced amdS as sole source of nitrogen. Suitably, a medium for the method according to the invention comprises a carbon source other than glucose. Suitably, a medium for use in a method according to the invention comprises sorbitol as carbon source, preferably as sole carbon source. A suitable medium at least further comprises minimum substrates required for growth of *A. sojae*.

A method according to the invention as defined above in any of the embodiments comprises introduction of an additional nucleic acid sequence besides the *amdS* gene. The additional nucleic acid sequence for example encodes a protein or polypeptide, such as a phytase or proteins having phytase activity. The sequence does not necessarily have to be a non *Aspergillus sojae* sequence, but can also include *A. sojae* derived sequences It is however intended to indicate that the sequence that is introduced is absent in the non-transformed strain or else is present in a lower copy number than in the *A. sojae* according to the invention.

Naturally, the subject invention also covers any *Aspergillus sojae* obtained by the method described above. Basically, the method is directed at introducing a sequence capable of realising the presence of sufficient active amdS to function as selectable marker as opposed to the *A. sojae* into which the sequence is introduced which cannot for some reason or another produce sufficient active amdS to enable growth on a substrate for amdS as sole source of nitrogen.

A method of selecting transformed or transfected *A. sojae* also falls within the scope of the invention. The method comprises subjecting *A. sojae* (with no active endogenous *amdS* gene as defined according to any of the embodiments described) to a method of transformation or transfection of the *A. sojae* in a manner known per se for transformation or transfection of fungi with a nucleic acid sequence. The method comprises the introduction of an *amdS* gene as the nucleic acid sequence, followed by selection of the resulting transformed or transfected *A. sojae,* said selection occurring on a medium comprising a substrate for the introduced amdS as sole source of nitrogen, said medium further comprising a carbon substrate, said medium enabling the desired *A. sojae* to grow whilst eliminating growth of non transformed or transfected *A. sojae* due to inability of such to grow without the introduced *amdS* gene on the selection medium. A suitable medium at least further comprises minimum substrates required for growth of *A. sojae*.

The invention is also directed at a method for producing recombinant *Aspergillus*

*sojae*. This method comprises introducing a desired nucleic acid sequence e.g. by transformation or transfection in a manner known per se into an *A. sojae*, said desired nucleic acid sequence being flanked by sections of the endogenous *amdS* gene of a length and homology sufficient to ensure recombination. The introduction is followed by selection of the recombinant *A. sojae* having the desired nucleic acid sequence. The selection occurs for a selectable marker comprised in or transformed in cotransformation with the desired nucleic acid sequence, said selectable marker being absent in the *A. sojae* prior to introduction of the desired nucleic acid sequence. The flanking sequences may also be sequences corresponding to the endogenous *amdS* gene sufficient to ensure recombination. The skilled person can readily assess which sequences will suffice on the basis of hybridisation knowledge and the sequence data of the endogenous *amdS* gene. The recombination event eliminates the endogenous amdS activity in both cases. The selectable marker can quite suitably be pyrG, with, however, uracil instead of uridin in the selection medium.

A further embodiment of the invention comprises *Aspergillus sojae* exhibiting growth on medium comprising uracil and fluoro-orotic acid, said *A. sojae* further not exhibiting growth on medium comprising uridine and fluoro-orotic acid. This means that the *A. sojae* exhibits uracil auxotrophy, is unable to utilize uridine, is pyrG negative and exhibits resistance to fluoro-orotic acid. The uracil auxotrophy and the fluoro-orotic acid resistance are relievable upon complementation with an active introduced *pyrG* gene. Such an *A. sojae* according to the invention can be free of active endogenous *pyrG* genes. The pyrG negative *A. sojae* according to the invention may comprise an endogenous *pyrG* gene with a mutation inactivating it. The mutation can be any mutation known or conceivable to a person skilled in the art, said mutation inactivating a *pyrG* gene or the expression product thereof. Such a mutation can by way of example be in the form of an insert of a nucleic acid sequence in the gene, a substitution of a part of the encoding sequence of the gene, a deletion of a part of the encoding sequence of the gene or a deletion of the whole encoding sequence of the gene. The mutation can also occur in the regulating part of the gene. In the case of *Aspergillus sojae* according to the invention with a mutated *pyrG* gene, said *Aspergillus sojae* can have a nucleic acid sequence for the mutated *pyrG* gene different to that of the wild type *A. sojae pyrG* gene. A further embodiment comprises pyrG negative *A. sojae* according to the invention as described in any of the above embodiments which further comprise any of the characteristics described for any of the amdS variant *A. sojae* according to the invention as such or in combination.

A method of selecting transformed or transfected *Aspergillus sojae* also falls within the scope of the invention. The method comprises subjecting *A. sojae* of the pyrG negative type according to any of the embodiments of the invention as described above to a method of transformation or transfection with a nucleic acid sequence, said method comprising introducing an active *pyrG* gene into the pyrG negative *A. sojae* in a manner known per se for transformation or transfection. The introduction step is then followed by selection of the resulting transformed or transfected *A. sojae* on a medium free of uracil and fluoro-orotic acid, said medium at least further comprising minimum substrates required for growth of *A. sojae*, said medium enabling the desired *A. sojae* to grow whilst eliminating growth of non-transformed or -transfected *A. sojae* due to inability of such to grow without uracil due to the inactivated *pyrG* gene. In a suitable embodiment of such a method the active *pyrG* gene that is introduced is flanked by identical nucleic acid sequence fragments, and the pyrG positive *A. sojae* resulting from the introduction of the *pyrG* gene and the flanking sequences is selected on a medium free of uracil and fluoro-orotic acid. Subsequently the pyrG positive *A. sojae* is cultivated on medium comprising uracil and fluoro-orotic acid, thereby eliminating the *pyrG* gene that had been introduced and thus resulting in a pyrG negative *A. sojae* that is selectable by growth on uracil comprising medium and fluoro-orotic acid resistance. In a suitable embodiment of the aforementioned method the flanking sequences and the *pyrG* gene are further flanked by sequences that direct integration of the *pyrG* gene and the flanking sequences into a specific location, due to the fact that the integration directing sequences are homologous to a specific sequence of the *A. sojae* to be transformed. This enables knock out (if desired) of the gene associated with the specific sequence. The process of knock-out mutant creation as such is well known to the person skilled in the art. Any of the embodiments of the selection method just described may further comprise the step wherein the *Aspergillus sojae* is transformed or transfected with a further heterologous nucleic acid sequence. The further heterologous nucleic acid sequence preferably encodes a protein or polypeptide and the same remarks are valid here as made elsewhere in this description for the nature of such further nucleic acid sequences for the other embodiments of *Aspergillus sojae* and fungi in general according to the invention. The further sequence can be introduced with the active *pyrG* gene either on the same vector or by cotransformation with the active *pyrG* gene that is introduced. The method of selecting transformed or transfected *A. sojae* as described may also be carried out in combination with the method for introducing a nucleic acid comprising introduction of a heterologous *amdS* gene in any of the embodiments according

to the invention disclosed therefore above. Naturally, the invention covers any recombinant *A. sojae* obtained by the method of selecting transformed or transfected *A. sojae* according to the invention.

The invention is also directed at a method for producing recombinant *Aspergillus sojae*, said method comprising transformation or transfection in a manner known per se of a pyrG positive *Aspergillus sojae* with a nucleic acid sequence comprising the sequence to be introduced flanked by sections of the *pyrG* gene or corresponding sequences of a length and homology sufficient to ensure recombination eliminating the *pyrG* gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for the *A sojae* with a pyrG negative phenotype. Determination of the corresponding sequences lies within the reach of the skilled person by virtue of their knowledge of hybridisation processes with nucleic acid sequences and their knowledge of required sequence data of the *pyrG* genes.

In particular the invention also covers such *Aspergillus sojae* exhibiting the characteristics of the amdS variant *A sojae* according to the invention as defined above. Thus any *Aspergillus sojae* strain obtained by either the amdS and/or pyrG introduction method according to the invention is a novel strain falling within the scope of the invention as is any subsequent use of such a novel strain. Such a novel strain can comprise nucleic acid sequences that do not occur in the original corresponding *Aspergillus sojae* strain or even do not occur in *Aspergillus sojae, Aspergilli* or fungi. The sequences can be of mammalian origin or derived from any animal, plant or microbe. Nucleic acid sequences can also be expressed that are naturally present in the *Aspergillus sojae* strain but that are present in a lower copy number in the corresponding non-transformed *A sojae*. Thus the production of homologous proteins is also covered by the invention when pyrG and/or amdS *Aspergillus sojae* strains according to the invention are involved. A preferred embodiment is that wherein the particular protein or polypeptide to be produced is absent in the corresponding non-treated *A. sojae* and/or is present in a lower copy number in the corresponding non-treated *A. sojae*, i.e. the *A. sojae* prior to introduction of the nucleic acid sequence. Expression of heterologous proteins by any of the novel strains of *Aspergillus sojae* in a manner known per se for producing protein or polypeptide in a fungus thus covers both sequence native to the strain and foreign to the strain. Basically, only the native non-transformed or -transfected *A sojae* is excluded from protection. A process of production comprises cultivating the fungus under suitable conditions for expression of the desired sequence to occur. The process of production optionally includes

the step of isolation of the resulting polypeptide or protein in a manner known per se for protein or polypeptide production by fungi. Preferably the protein or polypeptide will be secreted into the culture medium.

A preferred protein or polypeptide is a protein or polypeptide susceptible to degradation upon expression by *Aspergillus niger* or *Aspergillus awamori*. A number of such proteins and polypeptides have already been disclosed in the prior art and a large number remain yet to be determined. Such determination is however a matter of routine for the skilled person. Another preferred embodiment of the protein or polypeptide to be expressed is one whereby the protein or polypeptide differs from an *Aspergillus sojae* protease and amylase. A preferred embodiment involves a non *Aspergillus sojae* protein or polypeptide.

A particularly interesting embodiment comprises a combination of the two processes for introducing nucleic acid sequences according to the invention as described above. The advantage thereof lies in the fact that the frequency of transformation obtained with the pyrG marker is clearly much higher than that of the amdS marker. However, secondary screening of the pyrG+ strains for the best growth on acrylamide selective plates allows the identification of those recombinant *Aspergillus sojae* showing the highest copy number and thus most likely the highest level of gene expression.

As indicated in the examples homologous and heterologous expression regulating sequences can be used by *Aspergillus sojae* i.e. natively occurring sequences of the strain itself or sequences foreign to the strain can be used. Thus the transformants according to the invention can comprise any such regulatory sequences. The selection of the suitable regulatory region is a matter of choice that lies well within the range of the standard capabilities of the skilled person and will depend on the particular application. The regulating sequences can be constitutive or inducible. The regulating sequences can be fungal or non-fungal. A broad range are exemplified in the examples. A large number of expression regulating sequences are regularly used in the art for other systems, in particular fungal systems such as *Aspergilli*, and can routinely be applied without undue burden in the *Aspergilli* according to the invention.

For introducing the desired nucleic acid sequences into *Aspergillus sojae* any vector may be used that is suitable for introducing nucleic acid sequences into fungal host cells. Numerous examples are available in the art. In particular vectors that have been found suitable for transformation, transfection or expression in *Aspergilli* such as *Aspergillus niger*, *Aspergillus awamori* and *Aspergillus oryzae* can suitably be applied.

In addition to the above the subject invention describes efficient protein production for recombinant *Aspergillus sojae*. Such efficient production is disclosed in those strains having a protease profile superior to ATCC42251 or at least as good as any of ATCC9362, ATCC11906 and ATCC20387. The subject description thus reveals that some known strains of *A. sojae* are well suited already as such for production of proteins, polypeptides and metabolites. These *Aspergillus sojae* strains exhibit a lower proteolytic activity than the reference strain *A. sojae* ATCC42251. In particular the two known strains ATCC11906 and ATCC20387 are well suited. So preferred *A. sojae* strains for production of proteins, polypeptides and metabolites will be those expressing equal to or less proteolytic activity than the two preferred strains. Strain ATCC11906 is the best embodiment of the deposited ATCC *A. sojae* strains according to the prior art. Suitable proteins or polypeptides will be produced. Now that the subject invention has enabled introduction of nucleic acid sequences, such can serve to provide any protein or polypeptide of choice using an *A. sojae* as expression host.

The subject invention offers an improvement over existing expression systems. A number of existing protein production systems exhibit expression problems due to proteolysis. In particular the new system is better than the currently frequently applied expression systems *Aspergillus niger* and *Aspergillus awamori*. The subject invention now renders it possible to provide a recombinant *Aspergillus sojae* comprising a introduced nucleic acid sequence encoding a protein or a polypeptide for expression, said protein or polypeptide being susceptible to degradation upon expression by *A. niger* or *A. awamori*. The invention also provides a recombinant *A. sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being other than *A. sojae* protease and amylase. A preferred embodiment is that wherein the introduced nucleic acid sequence encodes a non-*A. sojae* protein or polypeptide. Such recombinant *A. sojae* strains also fall within the scope of the invention.

In addition, illustration of *Aspergillus sojae* strains that have been modified in order to enhance their suitability as expression hosts is currently provided. These modifications can be reduced proteolytic activity as induced by any means. Specifically, the use of UV random mutagenesis is illustrated. Also specific mutation of one or more protease genes is illustrated. The means by which mutations can be introduced are common knowledge to the skilled person, and numerous alternative embodiments are thus readily available to arrive at the desired mutants. A suitable embodiment is formed by mutants in which alkaline proteolytic activity has been reduced. In particular elimination of activity of

specifically the major 35 kDa alkaline protease is illustrated as ensuring increased expression of proteins and polypeptides. Specifically the invention thus also covers novel strains exhibiting reduced proteolytic activity, specifically reduced alkaline proteolytic activity. Such strains are obtainable using any specific mutation route known or conceivable to the skilled person. A preferred embodiment of such expression hosts exhibiting reduced proteolytic activity as described above further comprises a selectable marker. Quite suitably the selectable marker will be amdS, pyrG or a combination thereof.

The invention in particular covers a method of producing protease deficient mutants of *A. sojae* by knocking out the 35 kDa alkaline protease gene. There are numerous ways in which this can potentially be achieved on the basis of the sequence data provided for this gene. In particular a method using recombination with a pyrG selection marker linked to two flanking regions eliciting cross over of the 35 kDa alkaline protease gene, whereby the resulting strain has the pyrG selection marker and misses the 35 kDa alkaline protease gene is an elegant one. Subsequently the pyrG selection marker can be eliminated, thus providing a 35 kDa alkaline protease negative *Aspergillus sojae* mutant that can be used for expression purposes of any desired sequence to be introduced therein. Naturally, the sequence to be introduced can have been incorporated in the previous steps already either on the same vector as the pyrG marker or in a cotransformation event. Also the method can be carried out analogously where a different protease gene than the 35 kDa alkaline protease gene is to be knocked out. The analogous measures to be taken are obvious to the skilled person on the basis of the illustration provided herein in combination with knowledge of other protease sequences. Also analogously the amdS selectable marker can be used in accordance with the invention as described elsewhere in this description.

Mutant fungi exhibiting improved fermentation characteristics are also provided as an additional aspect of the subject invention. Specifically, the invention is directed at a fungus comprising a mutation inhibiting the activity of proprotein convertase or an equivalent protein. Numerous proprotein convertases are known in the art. In particular we refer to Figure 1 providing sequence data of a number of such proteins. A fungus according to the invention is suitably selected from *Agaricus*, *Aspergillus*, *Trichoderma*, *Rhizopus*, *Mucor*, *Phanerochaete*, *Trametes*, *Penicillium*, *Cephalosporium*, *Neurospora*, *Tolypocladium* and *Thielavia*. Particularly suitable fungi are *Aspergillus niger*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus awamori*, *Aspergillus oryzae*, *Trichoderma reesei*, *Penicillium chrysosporum*, *Cephalosporium acremonium*, *Neurospora crassa*, *Tolypocladium geodes and Thielavia terrestris*. A preferred embodiment covers the mutant

when it is an *Aspergillus sojae,* most particular preference is extended to *Aspergillus sojae* as defined above according to the invention, i.e. comprising heterologous nucleic acid sequences, e.g. in combination with the selectable markers amdS and/or pyrG.

A suitable equivalent of a proprotein convertase is a protein or polypeptide exhibiting an amino acid sequence with more than 40%, preferably more than 45% similarity or identity with the inferred amino acid sequences of the DNA sequences given in SEQ ID No. 3 (= gene fragment encoding *A. niger* proprotein convertase amino acid sequence), SEQ ID No. 4 (= partial gene fragment encoding *Aspergillus sojae* proprotein convertase amino acid sequence) or with any of the sequences given in Seq ID Nos. 5 to 9. The functionally equivalent protein may suitably have a nucleic acid sequence capable of hybridising under stringent conditions to a nucleic acid sequence according to SEQ ID Nos. 3 to 9. Stringent hybridisation conditions can readily be determined by the skilled person. A suitable example of stringent hybridisation conditions are hybridisation at 50°C and preferably at 56°C and final washes at 3xSCC. *PE4*, *PCL1* and *PCL2* are specifically mentioned as examples of suitable oligonucleotide mixtures corresponding to the coding strand (i.e. SEQ ID Nos. 10, 11 and 12). For the noncoding strand *PE6*, *PCL2-rev*, *PCL3* and *PCL4* are mentioned (i.e. SEQ ID Nos. 13, 14, 15 and 16, respectively). Use of these primers in amplification procedures common in the art will provide equivalent sequences and such use and the resulting newly found sequences and application thereof in the manner analogous to that described in the subject description fall within the scope of the invention. The sequences for which the oligonucleotides were made are well conserved as could be determined from comparison of the various amino acid sequences for the proteins provided (see Figure 1). Any other nucleic acid sequences exhibiting the same or higher degree of identity, similarity or homology with the sequences provided in the subject patent application for the proteins or relevant active parts thereof are covered by the invention as is the use thereof as primers or probes to find other proprotein convertase or equivalent protein encoding sequences and/or for subsequently introducing mutations in such protein encoding sequences. By way of example Maniatis *et al.* (1982) Molecular Cloning, A Laboratory manual, Cold Spring Harbor Laboratory, New York or any other handbook on cloning and/or screening nucleic acid sequences has been referred to. The equivalent protein or polypeptide will exhibit the activity of a proprotein convertase as the one having an amino acid sequence according to SEQ ID Nos. 3 to 9. The mutant fungus can comprise a substitution, insertion or deletion in the encoding sequence of the proprotein convertase or equivalent protein. The mutant fungus can suitably comprise a

mutation in the regulation of the expression of the gene encoding proprotein convertase or equivalent protein. A mutant fungus according to the invention in a suitable embodiment exhibits reduced viscosity vis a vis the corresponding non mutated fungus under equivalent cultivation conditions. A mutant fungus according to any of the above embodiments exhibiting increased expression of a desired introduced nucleic acid sequence encoding a protein or polypeptide is included within the scope of the invention, said fungus exhibiting increased production of a protein or polypeptide under equivalent conditions vis a vis the corresponding wild type fungus. The activity site for the *A. sojae* proprotein convertase has been ascertained to be comprised within the amino acid sequences inferred by SEQ ID Nos. 3 and 4.

A process for producing a phytase or protein having phytase activity or any other protein or polypeptide, preferably a recombinant phytase or any other heterologous protein or any other polypeptide, said process comprising cultivating a mutant fungus according to any of the embodiments described above falls within the scope of the invention. A process for obtaining the resulting protein or polypeptide either from the cell as such or after secretion thereof from the cell is also included.

The use of any of the described novel strains for transformation of any nucleic acid sequence encoding a phytase or protein having phytase activity or any other protein or polypeptide thereto and any subsequent expression of any nucleic acid sequence introduced therein and also optionally any following processing and/or secretion and/or isolation is covered by the invention.

Any phytase or phytase-like or any other heterologous protein or polypeptide encoding sequence can suitably be used. This can be of fungal or non-fungal origin. A preferred embodiment is formed by acid labile protein or polypeptide encoding sequences. Suitably the protein encoding sequence encodes non protease-like proteins. The examples show a phytase sequence and a number of heterologous sequences suitable for use in transformation and also for expression in *Aspergillus sojae* hosts. Further examples of suitable proteins to be expressed are obvious to a person skilled in the art.

The invention is further illustrated by the examples below. The examples are not to be considered restrictive to the interpretation of the scope of the invention. Alternative embodiments are readily envisageable to the skilled person on the basis of the description and knowledge of the relevant field of technology. The content of references cited in the description are incorporated by reference. The claims serve to illustrate the intended scope of the invention.

EXPERIMENTAL DETAILS CONCERNING THE INVENTION

CONSTRUCTION OF AN *Aspergillus sojae* GENE LIBRARY.

Genomic DNA of *A. sojae* was isolated from protoplasts obtained from ATCC11906 using a previously described protocol (Punt, van den Hondel, 1992). After isolation, DNA was extracted from the protoplasts using the protocol described by *Kolar et al.*, 1988. Subsequently the DNA was partially digested with *Mbo*I to result in DNA fragments of an average size of 30-50 kb.

Vector pAOpyrGcosarp1, which was used for the construction of the gene library, was constructed by ligation of a 3 kb *Bam*HI-*Hind*III fragment from pANsCos1 (Osiewacs, 1994) and a 3.2 kb *Acc*65I-*Hind*III fragment from pAO4.2 (De Ruiter-Jacobs, 1989) in *Acc*65I-*Bam*HI digested pHELP1 (Gems et al., 1991). This cosmid vector carries the *A. oryzae* pyrG selection marker and is self-replicating in filamentous fungi.
*Mbo*I digested genomic DNA was ligated to *Bam*HI-digested pAOpyrGcosarp1, and the ligation mixture was packaged into phage particles using the Stratagene Supercos1 vector kit. In total 30,000 individual clones were obtained representing an approximate 30-fold representation of the *A. sojae* genome. Stocks (in 15% glycerol) of pools of the resulting clones were stored at -80°C for later use.

AMDS TRANSFORMATION METHOD AND TRANSFORMANTS.

Two currently used protoplasting protocols and transformation protocols [the modified OM-method (Yelton et al., P.N.A.S. 81 (1984) 1470-1474) and the NaCl-method (Punt and Van den Hondel, Meth. Enzym. 216 (1993) 447-457)] were tested on the *Aspergillus sojae* strain ATCC9362. Both methods resulted in protoplasts, but the yield of viable protoplasts with the OM-method was clearly better. The overall yields were lower than normally obtained for *A. niger*. A pilot protoplasting/transformation experiment was carried out with all *A. sojae* strains using the OM method.

For transformation, vector p3SR2 (carrying the amdS marker) was used in combination with pAOpyrGcosARP1. This latter vector is a derivative of the autonomously replicating *Aspergillus* vector Arp1, which in all *Aspergillus* species tested so far, resulted in highly increased numbers of (instable) transformants when used as a cotransforming vector. For nearly all strains sufficient protoplasts (about 10E6-10E7 per transformation) were obtained. Analysis of appropriate AmdS selection conditions for the various *A. sojae* strains revealed vigorous growth of most strains on the commonly used

selective acetamide medium. Clearly, the acetamide selection conditions proposed for *A. sojae* amdS transformants as reported in WO97/04108, were not appropriate for the selection of *A. sojae* transformants. Our experiments revealed, surprisingly, that AmdS+ transformants could only be selected with acrylamide selection. Even on selective acrylamide plates, a considerable background from non-transformed protoplasts was observed. Selection of primary transformants requires around three weeks and many of the initially selected putative transformants turned out to be false positives, only showing background growth after transfer to fresh selective acrylamide plates. To optimize selection of transformants attempts had to be made to reduce this background growth. Improved results were obtained by omitting glucose from the selective plates. In Table 3 the composition of the improved selection medium and the usual media is given. Figures 2a, b and c show the background growth observed for selected strains on the selection medium described in WO97/04108 and the improved acrylamide selection medium described in Table 3.

Further transformation experiments with the three selected *A. sojae* strains revealed that protoplasting efficiencies for ATCC11906 and ATCC20387 were better using the NaCl-method. Successful protoplasting was obtained using various commercially available protoplasting enzyme preparations such as NOVOZYM, Caylase, Glucanex, etc. Based on the NaCl transformation protocol the three selected *A. sojae* strains were transformed with amdS selection vector p3SR2 or derivatives thereof. Using the modified acrylamide selection plates a number of vigorously growing transformants were obtained, while no growth was observed in a control transformation without DNA. Another approach to circumvent background growth of non-transformed mycelium is the elimination of the activity of the wild type *A. sojae amdS* gene. This can be achieved for example by disruption of the *A. sojae amdS* gene. As a first step specific DNA fragments carrying ATCC11906 *amdS* sequences were PCR-amplified using primers derived from published *A. oryzae amdS* sequences (Gomi *et al.*; 1991, Gene 108, 91-98). Previous experiments had shown that cloning by stringent hybridisation would be unsuccessful due to a low level of sequence conservation between *A. nidulans* and *A. sojae amdS* sequences. The expected fragment of about 1.6 kb, which should carry most of the coding region of the *amdS* gene, was obtained. Sequence analysis from both ends of the cloned PCR fragment (Figures 3a and 3b) confirmed the cloning of a part of the *A. sojae amdS* gene. The stringent hybridisation occurred at 56°C with final washes at 3xSSC. The cloned sequence was very similar to the published *A. oryzae amdS* sequence. Several hybridising clones (7 out of

10.000) were isolated from the ATCC11906 cosmid library in pAOpyrGcosarp1 using the cloned ATCC11906 *amdS* fragment as a probe. After subcloning a fragment carrying the complete *amdS* gene, a part of the *amdS* gene was replaced by a re-usable pyrG selection marker to generate an amdS replacement vector. Transformation of this vector to *Aspergillus sojae* ATCC11906PyrG resulted in pyrG+ transformants. After subsequent analysis of these transformants on acetamide and acrylamide selection plates several of these transformants showed reduced background growth. Southern analysis of a few of these strains revealed that the expected gene replacement had occurred. One of these strains was used for subsequent transformation with the *A. nidulans amdS* gene using acrylamide selection plates and resulted in a number of amdS+ transformants.

PYRG TRANSFORMATION METHOD AND TRANSFORMANTS.

*(1) Initial experiments*

For *A. sojae*, the standard experiments used in the prior art for other fungi to generate pyrG mutants as described in the introduction resulted in numerous fluo-orotic acid (FOA)-resistant strains. However, all of these strains were able to grow on medium without uridine and were therefore not considered pyrG mutants. With our final aim to isolate the appropriate mutant strains a number of alternative approaches were followed.

*(2) Near-homologous gene disruption*

Based on the expectation that the *pyrG* genes from *A. sojae* and *A. oryzae* are very similar in sequence (which was confirmed by Southern hybridisation carried out under stringent conditions), experiments were carried out to disrupt the *A. sojae pyrG* gene with a mutant version of the *A. oryzae pyrG* gene using an approach previously described by Gouka et al. (1996). The stringent hybridisation occurred at 65°C with final washes at 0.3 x SSC. An *A. oryzae pyrG* disruption vector was constructed in which an 0.5 kb *Cla*I fragment carrying part of the *pyrG* coding region was deleted (Figure 4). The *Xba*I *pyrG* fragment from this new vector was used for transformation and direct selection for FOA resistant transformants. None of the FOA resistant colonies obtained was uridine requiring.

*(3) UV mutagenesis and filtration enrichment*

Another approach to improve the yield of specific mutant strains is the use of a filtration-enrichment step (Bos *et al.* 1986, Thesis, Agricultural University Wageningen). The UV mutagenized spores are used for inoculation of a minimal medium (MM) liquid

culture. From the resulting repeated overnight culture those spores unable to germinate in minimal medium (a.o. pyrG mutant spores) are separated from the grown mycelium by filtration through myracloth. The spores obtained after several enrichment steps were tested for their PyrG phenotype, by inoculating the spores on plates containing FOA. Again none of the resulting FOA resistant colonies was uridine requiring. Also none of the colonies obtained after this enrichment on MM plates containing uridine was shown to be uridine requiring.

*(4) Modified selection conditions*

Our previous attempts to isolate pyrG mutants from *A. sojae* had failed suggesting the inability of the required pyrG mutants to utilize exogenous uridine, which is used in the FOA selection medium for the analysis of uridine auxotrophy. A modified selective FOA medium, now containing uracil next to uridine, was used in a new isolation attempt. From this approach several FOA resistant mutants were obtained which were uracil requiring. Retesting of these strains showed that these were unable to grow on uridine supplemented minimal medium. Subsequent transformation experiments with some of the uracil-requiring strains showed that these mutants could indeed be complemented with a fungal *pyrG* gene (e.g. vector pAB4.1; *A. niger pyrG*). The inability of pyrG mutants to grow in minimal medium supplemented with only uridine was an unprecedented observation for related *Aspergillus* species (*A. nidulans, A. niger, A. oryzae*) and various other fungal species.

*(5) Re-usable selection marker*

Versatile genetic modification of *A. sojae* requires the possibility to modify, disrupt and express a number of different genes in a single fungal strain, which would require the availability of a (series) of different selection markers. However, the availability of a marker such as pyrG, which allows selection of both the mutant (FOA selection) and the transformant (Uracil-less medium), provides the possibility of repeated use of the same marker in subsequent experiments. For this approach a *pyrG* marker gene was designed, in which the complementing sequence was flanked by a direct repeat sequence originating from the 3' flanking end of the *pyrG* gene. The resulting plasmid is pAB4-1rep. The construction of this vector is detailed in Figure 5. The full sequence of the vector is given in SEQ ID No. 17. Transformation of *A. sojae* pyrG mutants with this vector results in a similar number of PyrG+ transformants as with the vector pAB4-1. However, subsequent

plating of spores of selected pAB4-1 and pAB4-1rep transformants to FOA selection plates resulted in many more FOA resistant/uracil requiring colonies for the pAB4-1rep transformant. Southern analysis of these FOA resistant/uracil requiring clones showed that in most of the pAB4-1rep strains the *A. niger pyrG* marker gene had been deleted leaving only the small 0.7 kb repeat region at the locus of integration, while in the pAB4-1 strains the *A. niger* gene was still present and had presumably acquired a mutation resulting in the pyrG-negative phenotype.

## EXPRESSION HOSTS: STRAIN SELECTION.

*Protease production*

Very important characteristics of a fungal expression system are the level and type of fungal proteases produced under various culture conditions. Sometimes strains which can be readily transformed are not suitable as expression hosts due to production of proteases or acidification of the culture media which is detrimental to the expressed product. Analysis of the growth behaviour of the various *A. sojae* strains revealed that, in contrast to what was observed for *A. niger*, acidification of the culture medium did not occur either on agar based plates (MacConkey) or in shake flask cultures. In fact in shake flask cultures, irrespective of the three medium types analyzed (Table 4), in most cases even an alkaline pH was obtained in the cultures. Based on these results and literature data it is thus expected that primarily alkaline proteases will be present in the *A. sojae* culture fluid. To analyse protease activity of the culture fluids of the various strains, a milk clearing assay was performed. In addition medium samples were incubated with different proteins (e.g. bovine serum albumin (BSA)), and degradation of these proteins was followed in time in order to assess the suitability of the tested strains as expression hosts for a range of products. BSA was chosen as in our previous experiments with *A. niger*. This protein was shown to be very susceptible to proteases. *A. terreus* phytase was chosen as example of another proteolytically instable protein. Degradation of milk proteins as shown by the formation of a milk clearing zone at the periphery of growing colonies is a generally accepted criterion for protease activity. Detection of BSA was carried out by Coomassie staining of SDS-PAGE gels. For phytase, Western analysis using specific antibodies, was carried out. As shown in Table 4, clear differences of degradation in *A. niger* culture fluid are evident when this is compared with that in *A. sojae* culture fluid. In *A. niger* culture fluid (pH 3-4) rapid degradation of BSA occurs. In *A. sojae* culture fluids

from richer media, degradation of BSA occurs, albeit less than in *A. niger* culture fluid. In most *A. sojae* culture fluids (pH 7-8) rapid degradation of *A. terreus* phytase occurs, with the exception of ATCC9362, ATCC11906 and ATCC20387 culture fluids. In general, the strains with the lowest phytase degradation also show low BSA degradation under the conditions tested. In particular the two *A. oryzae* strains ATCC20235 and ATCC46250 show much higher proteolytic activity than most *A. sojae* strains.

To exclude that differences in the pH of the culture fluid cause the observed effects, similar degradation experiments were also carried out with culture fluids of which the pH was adjusted to pH 4.5 (50 mM Na/HAc), pH 5.8 (50 mM Na/HAc) and pH 8.3 (50 mM Tris/HCl). Table 5 gives the degradation data obtained with these samples. As can be seen in the table *A. oryzae* ATCC20235, which had the highest proteolytic activity at pH 7-8 also shows high proteolysis at other pH values. Degradation of *A. terreus* phytase occurs primarily at pH 8. Similarly to what was found before, ATCC11906 and ATCC20387 showed low phytase degrading activity. ATCC9362 showed phytase degradation in rich media. BSA degradation by *A. sojae* showed no significant differences with the data presented in Table 4.

In conclusion, these protease assays resulted in the identification of three low protease *A. sojae* strains, namely ATCC9362, ATCC11906 and ATCC20387. Thus, *A. sojae* can cleary be used as expression host for a range of proteins and provides a series of advantages over prior art transformation and expression systems.

STRAIN IMPROVEMENT

Once the potential of transformability and expression had been ascertained for *Aspergillus sojae*, means by which additional strains could be created with enhanced characteristics for expression were considered. Two different approaches which can be used as such or in combination were developed to provide novel improved strains for expression of proteins.

On the one hand the possibility of developing protease deficient mutants was investigated and the impact of such on levels of expression was assessed. On the other hand strains with amended morphology were developed with a view to improve fermentation characteristics. To achieve this a hitherto non-disclosed or suggested route was followed which is applicable not only to *Aspergillus sojae* but to *Aspergilli* and in fact to fungi in general.

*Development of protease deficient mutants*

To obtain protease deficient *A. sojae* strains two approaches were followed. In a first approach spores from ATCC11906 and ATCC11906-derived strains were mutagenized with UV. In a second approach gene disruption of the major alkaline protease was carried out.

*UV mutants*

Freshly harvested spores from *A. sojae* ATCC11906 or one of its pyrG derivatives was UV-mutagenized in a Biorad UV-chamber with a dose resulting in 20-50 % survival. Serial dilutions were plated onto skim-milk plates (Mattern et al., 1992). From 5000 UV-surviving strains four mutant strains with a considerably reduced milk-clearing halo were obtained

*AlpA gene disruption*

In this approach the endogenous *alpA* (alkaline protease) gene cloned from ATCC11906 was disrupted using a disruption vector carrying the re-usable pyrG selection marker as described in this description.

An ATCC11906 cosmid library (in a PyrG cosmid) was constructed. From 10.000 independent cosmid clones initially 4 were found to hybridize under homologous conditions with an *A. sojae alpA* fragment obtained by PCR with primers MBL1784 and MBL1785. Rescreening of the 4 clones revealed only strong hybridisation with one clone. A 4 kb *Eco*RI and a 2.5 kb *Hind*III fragment from this clone, together expected to carry the complete gene, were subcloned and characterised by restriction enzyme digestion and sequence analysis. Based on these subclones a new gene-replacement vector was constructed as described in Figure 6. For transformation of an ATCC11906pyrG derivative the vector was digested with *Eco*RI, and the 8.7 kb *alpA* deletion fragment was used for transformation (see Figure 6). Transformation of the replacement cassette to ATCC11906pyrG5 resulted in a number of transformants with a reduced milk-halo. Southern analysis of these strains revealed the successful deletion of the *alpA* gene. To allow subsequent use of the pyrG marker for transformation of one of these strains, spores from this strain were plated on FOA containing medium selective for pyrG mutants. From strains with the correctly integrated disruption cassette with the re-usable pyrG marker a large number of FOA resistant colonies were obtained. In contrast to the results obtained for spontaneous FOA resistant mutants of wild type strains, the FOA strains obtained from

these disruption strains were virtually all uracil requiring and turned out to be PyrG negative again. Southern analysis was used to confirm the desired removal of the *pyrG* marker gene at the *alpA* locus, leaving only the 700 bp "footprint".

*Analysis of protease activity in UV and disruption mutants*

To analyse the levels of protease production in the different low protease derivatives of ATCC11906 controlled batch fermentation experiments were carried out. From the resulting culture supernatants protease activities were determined at various pH values. Deletion of the *alpA* gene resulted in a strong reduction of proteolytic activity at alkaline pH. Analysis of the protease activity in one of the UV mutants showed almost complete absence of proteolytic activity at both pH 6 and pH 8. Consequently the level of proteolysis towards secreted proteins produced in these strains was considerably less than observed for the parent strain.

*Development of low viscosity mutants*

Initial controlled batch or fed batch fermentation trials with *A. sojae* resulted in considerable biomass yield, however both the culture viscosity and sporulation phenomena in the fermenter vessel represented less favourable characteristics.

Therefore attempts were made to improve these characteristics in the desired host strain. Various patent applications teach that low viscosity mutants can be isolated by various ways of screening. WO96/02653 and WO97/26330 describe non defined mutants exhibiting low viscosity. However here we describe a new unexpected case of a completely characterised and fully defined low viscosity mutant from *A. sojae*. It was found that a proprotein processing mutant from this organism had an unexpected aberrant growth phenotype (hyper-branching) while no detrimental effect on the productivity of proteins was observed. Controlled fermentation experiments with this strain revealed that increased biomass concentrations were obtained at considerably lower viscosity values. The observed characteristics were not only present in *A. sojae* but other fungi as well, e.g. in *A. niger*.

*(1) Construction of an A. niger proprotein processing mutant*

To clone the proprotein convertase encoding gene from *A. niger*, heterologous hybridisation using specific probes from the *Saccharomyces cerevisiae KEX2*, *Schizosaccharomyces pombe KEX1* and the *Xenopus laevis PC2* genes was carried out.

However, no specific hybridisation signals were obtained even at very low stringency hybridisation conditions (47°C, washes at 6xSSC), precluding the use of this approach to clone the corresponding *A. niger* gene.

As an alternative approach to clone the proprotein convertase encoding gene from *A. niger,* PCR was used. Based on the comparison of various proprotein convertase genes from various yeast species and higher eukaryotes (Figure 1) different PCR primers were designed (SEQ ID Nos. 10, 13 and 18-23) which are degenerated, respectively, 2048, 49152, 4, 2, 2, 512, 2048, and 4608 times. From the amplification using primers *PE4* and *PE6*, two individual clones were obtained of which the encoded protein sequence did show significant homology to the *S. cerevisiae KEX2* sequence (SEQ ID No. 24). These clones were used for further experiments.

Based on the observed homology to other proprotein convertase genes of the cloned PCR fragment, the corresponding *A. niger* gene was designated *pclA* (from *p*roprotein convertase-*l*ike). Southern analysis of genomic digests of *A. niger* revealed that the *pclA* gene was a single copy gene with no closely related genes in the *A. niger* genome, as even at stringent hybridisation conditions (50°C; washes at 6xSSC), no additional hybridisation signals were evident. A first screening of an EMBL3 genomic library of *A. niger* N401 (van Hartingsveldt *et al.*, 1987) did not result in any positively hybridising plaques although about 10-20 genome equivalents were screened. In a second screening a full length genomic copy of the *pclA* gene was isolated from an *A. niger* N400 genomic library in EMBL4 (Goosen *et al.*, 1987). Of the 8 hybridising plaques which were obtained after screening 5-10 genome equivalents, 6 were left after a first rescreening. All these 6 clones most likely carried a full copy of the *pclA* gene, as in all clones (as was observed for the genomic DNA) with the PCR fragment two hybridising *Eco*RV fragments of 3 and 4 kb were present (note that the PCR fragment (SEQ ID No. 24) contains an *Eco*RV restriction site). Based on a comparison of the size of other proprotein convertases, together these fragments will contain the complete *pclA* gene with 5'- and 3'-flanking sequences. The two *Eco*RV fragments and an overlapping 5 kb *Eco*RI fragment were subcloned for further characterisation. A detailed restriction map of the DNA fragment carrying the *pclA* gene is given in Figure 7.

Based on the restriction map given in Figure 7 the complete DNA sequence of the *pclA* gene was determined from the *Eco*RI and *Eco*RV subclones (SEQ ID No. 3). Analysis of the obtained sequence revealed an open reading frame with considerable similarity to that of the *S. cerevisiae KEX2* gene and other proprotein convertases. Based

on further comparison two putative intron sequences (SEQ ID No. 3, from position 1838 to 1889 and from 2132 to 2181) were identified in the coding region. Subsequent PCR analysis with primers flanking the putative introns, on a pEMBLyex based *A. niger* cDNA library revealed that only the most 5' of these two sequences represented an actual intron. The general structure of the encoded PclA protein was clearly similar to that of other proprotein convertases (SEQ ID No. 25 and Figure 8). The overall similarity of the PclA protein with the other proprotein convertases was about 50% (Figure 1).

To demonstrate that the cloned *pclA* gene is a functional gene encoding a functional protein, the construction of strains devoid of the *pclA* gene was attempted. Therefore, pPCL1A, a *pclA* deletion vector, in which a large part of the *pclA* coding region was replaced with the *A. oryzae pyrG* selection marker gene, was generated. Subsequently, from this vector the 5 kb *Eco*RI insert fragment was used for transformation of various *A. niger* strains. From these transformations (based on pyrG selection) numerous transformants were obtained. Interestingly, a fraction of the transformants (varying from 1-50%) displayed a very distinct aberrant phenotype (Figure 9). Southern analysis of several wild-type and aberrant transformants revealed that these aberrant transformants, which displayed a severely restricted growth phenotype, had lost the *pclA* gene. All strains displaying wild-type growth were shown to carry a copy of the replacement fragment integrated adjacent to the wild-type *pclA* gene or at a non-homologous position.

Analysis of protein production in selected pclA mutant strains carrying various glucoamylase fusion genes revealed the production of unprocessed fusion protein. The production of high levels of unprocessed glucoamylase-interleukin-6 fusion protein in a pclA mutant was achieved. Protein analysis revealed that in pclA mutant strains also no fully processed endogenous glucoamylase is formed but only pro-glucoamylase.

To further improve the yields of fusion proteins controlled batch and fed-batch fermentations were also carried out. Surprisingly the fermentation characteristics of pclA mutant strains were clearly superior to those of the parent strain resulting in a much reduced viscosity/biomass ratio, without loss of productivity.

*(2) Construction of an A. sojae proprotein processing mutant*

To construct the corresponding mutant in *A. sojae*, functional complementation of the low-viscosity mutant of *A. niger* genomic cosmid clones from the ATCC11906 cosmid library were isolated, which comprised the *A. sojae* proprotein processing protease *pclA* gene. Partial sequence analysis of the isolated sequences SEQ ID No. 4 confirmed the

cloning of the *A. sojae pclA* gene. Figure 10 shows the comparison of the DNA sequences of part of the *A. niger* and *A. sojae pclA* genes. Based on the *A. sojae* sequence and a partial restriction map with the coding region of the *A. sojae pclA* gene, a replacement vector was generated using the *Eco*RV-site in the *A. sojae pclA* gene to clone the re-usable *pyrG* marker as a *Sma*I fragment inside (Figure 11). The resulting vector was partially digested with *Cla*I to obtain the delta-pcl fragment of 10.5 kb (see Figure 11). This fragment was isolated to be used for transformation of *A.sojae* pyrG strains. The gene replacement vector was used to generate pclA mutants in ATCC11906 and ATCC11906 derivatives. The resulting strains were used for the expression of homologous and heterologous proteins. Controlled fermentation experiments with some of the resulting transformants revealed improved fermentation characteristics, in particular a lower viscosity/biomass ratio of the culture.

*(3) Cloning of fungal genes homologous to Aspergillus pclA*

Based on the comparison of amino acid sequences inferred from the *A. niger* and *A. sojae pclA* genes with those of other proprotein processing enzymes (Figure 1) several oligonucleotide mixtures corresponding to the coding or non-coding strand of well conserved sequences were designed (SEQ ID Nos. 10 to 16).

These oligonucleotide mixtures were used in PCR with chromosomal DNA from *Trichoderma reesei* QM9414, *Fusarium venenatum* ATCC20334, *Penicillium chrysogenum P2, Trametes versicolor, Rhizopus oryzae* ATCC200076, and *Agaricus bisporus* HORST. Depending on the template DNA used, PCR amplifications (30 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C) with one or more of the combinations of coding and non-coding strand oligonucleotides resulted in specific PCR products. Table 6 gives the results of the various amplification reactions. Sequence analysis was carried out with a number of the obtained PCR fragments using either of the two oligonucleotide mixtures used for amplification. These analyses resulted in the identification of the various *pclA* homologues from these different fungi. Figure 12 gives the inferred aminoacid sequences corresponding with the various DNA fragments (SEQ ID Nos. 5 to 9).

*(4) Examples of biomass and viscosity determinations*

The following operating parameter data ranges have been determined for fungal fermentations using a number of different fungal strains.

Viscosity:

Viscosity is determined on a Haake Viscotester VT500 using sensor system MV DIN (vessel number 7), operated at 20°C. Obtain a fresh sample of fermentation broth and place 40 ml of the broth in the measuring cell. After a small period of equilibration (4 min) at a set spindle speed the viscosity is measured. This measurement is repeated for ten different spindle speeds. Multiplication of the spindle speed with the measuring cell factor results in the shear rate. Viscosity $\eta$ (in centipoise = cP) is plotted against shear rate $\gamma$ (1/s) and gives the viscosity profile of the fermentation broth.

Viscosity ranges have been determined for fermentations using the specified fungal organism using the above procedure (Table 7).

Biomass:

Biomass is determined by the following procedure:

Preweigh 5.5 cm filter paper (Whatman 54) in an aluminium weighing dish. Filter 25.0 ml whole broth through the 5.5 cm paper on a Buchner funnel, wash the filter cake with 25.0 ml deionised water, place the washed cake and filter in a weighing pan and dry overnight at 60°C. Finish drying at 100°C for 1 hour, then place in desiccator to cool. Measure the weight of dried material. Total biomass (g/l) is equal to the difference between the initial and final weights multiplied by 40.

Protein:

Protein levels were determined using the BioRad Assay Procedure. The data presented above represent values determined 48 hours into the fermentation process until fermentation end; all values of *Aspergilli* and *Trichoderma* are for commercially relevant fungal organisms and reflect actual commercial data.

A fungal strain such as *A. sojae lfvA* and *A. sojae pclA* has the advantage that the low viscosity permits the use of lower power input and/or shear the in the fermentation to meet oxygen demands for those cases where shear stress on the product may be detrimental to productivity due to physical damage of the product molecule. The lower biomass production at high protein production indicates a more efficient organism in the conversion of fermentation media to product. Thus *A.sojae* mutants provides better biomass and viscosity data whilst also delivering at least as much protein, and in fact a lot more protein than the two commercially used systems which obviously are better than for typically deposited *Aspergillus* or *Trichoderma reesei* strains in general public collections.

The high protein production with low biomass concentration produced by *A. sojae lfvA* would allow development of fermentation conditions with higher multiples of increase in biomass, if increasing biomass results in increased productivity, for the desired product before reaching limiting fermentation conditions. The present high levels of biomass and viscosity produced by the *T. longibrachiatum* and *A. niger* organisms restrict the increase of biomass as the present levels of biomass and viscosity are near limiting practical fermentation conditions.

EFFICIENT GENE EXPRESSION

*(1) Heterologous regulatory sequences*

The three selected *A. sojae* strains were cotransformed with three GUS reporter vectors carrying different fungal expression signals (*A. nidulans* PgpdA; pGUS54, *A. niger* PglaA; pGUS64, *A. niger* PbipA; pBIPGUS) and the amdS selection vector p3SR2 or derivatives thereof. Expression of the GUS gene was analysed in representative transformants (Table 8). From the results it is clear that under the conditions tested the *gpdA* promoter was by far the best promoter resulting in about 5000 U GUS/mg protein. This corresponds to about 5% of the total amount of cellular protein. The *bipA* promoter results in about 30% of the gpdA activity, which corresponds to expression data obtained in *A. niger*. Surprisingly, the *glaA* promoter which is very active in *A. niger* (at least as active as gpdA) results in less than 1% of the gpdA activity in *A. sojae*.

*(2) A. sojae regulatory sequences*

We also isolated an *Aspergillus sojae* homologous promoter and assessed the applicability of such in an expression system. In some instances of expression it will be preferable to use a homologous promoter rather than a heterologous promoter. It was also interesting to assess whether the homologous promoter would be more efficient than a heterologous one.

PCR cloning of three efficiently expressed *A. sojae* genes, namely *alpA* (alkaline protease; inducible), *amyA* (amylase; inducible) and *gpdA* (glyceraldehyde-3-phosphate dehydrogenase; constitutive) was attempted using primers based on sequences available from *A. oryzae* (SEQ ID Nos. 26 to 31). Figures 13 a, b and c give the sequences and the position in published *A. oryzae* sequences of the various PCR primers used for this approach. Genomic template DNA from *A. sojae* ATCC11906 was used for PCR amplification. Initial PCR amplifications (30 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C) resulted in a specific PCR product of the expected size (400 bp) for the *gpdA*. For the other

two PCR reactions no product was obtained. Therefore, for *alpA* PCR conditions were made less stringent (10 cycles; 1 min 94°C; 1 min 25°C; 2 min 68°C + 20 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C), which resulted in a specific *alpA* PCR product of about 1000 bp.

The complete sequence of the cloned genes was determined. As shown in Figure 14 the *A. sojae* ATCC11906 *gpdA* promoter region has a very high homology with other *gpdA* promoter sequences and the *alpA* promoter was virtually identical to the *A. oryzae alpA* promoter (SEQ ID Nos. 32 and 33). Expression vectors carrying expression cassettes comprising these *A. sojae* promoters show significant levels of gene expression.

HETEROLOGOUS PROTEIN PRODUCTION

A number of heterologous proteins were tested which were known to be susceptible to acidic proteolysis and thus could not be expressed efficiently in other well known expression systems. Also proteins that are already efficiently expressed in alternative systems were tested in order to assess by way of comparison the levels of expression achieved with *Aspergillus sojae* vis a vis other known expression systems such as *Aspergillus niger*.

***Phytase production***

DNA fragments carrying various fungal phytases (Wyss et al. (1999) Appl. Environ. Microbiol. 65, 359-366) were ligated as 5' *Nco*I or *Bsp*HI sites introduced at the ATG codon - 3' blunt-ended fragments downstream of the *A.nidulans gpdA* promoter in pAN52-1NotI. The resulting vectors were used in cotransformation experiments of *A. sojae* using the *amdS* and or *pyrG* selection marker. Phytase producing transformants were screened using the described phytase plate-assay.

Further improved phytase expression vectors were generated using a multicopy cosmid approach. In this approach several copies of a phytase expression cassette, recloned in a multiple cloning site vector (pMTL series, Chambers *et al.*, (1988) Gene 68, 139-149) to allow its isolation as a *Eco*RI fragment. Several copies of these *Eco*RI fragments were cloned into cosmid vector pAN4cos1 through packaging (Verdoes *et al.* (1993) Transgenic Research 2, 84-92), resulting in cosmid clones carrying a number of expression cassettes. The resulting clones were introduced into *A. sojae* using the amdS selection marker. AmdS+ clones were screened for phytase production using the phytase plate-assay.

Further phytase expression vectors were generated using the GLA fusion approach (e.g. Broekhuijsen *et al.* (1993) J. Biotech. 31, 135-145). To this end phytase gene fragments, encoding the mature *A. fumigatus* phytase protein were fused, using convenient restriction sites and fusion PCR, to the 3'-end of the glucoamylase carrier gene in vector pAN56-1 (Genbank accession number Z32700). Between the glucoamylase and phytase part of the gene-fusion a sequence encoding a proprotein processing site (Asn-Val-Ile-Ser-Lys-Arg) was introduced. The resulting vectors were used in cotransformation experiments of *A. sojae* using the amdS and/or pyrG selection marker. Phytase producing transformants were screened using the described phytase plate-assay.

Shake flask fermentation was carried out resulting in significant levels of active phytase. Yield were significantly higher than those reported in literature for *A. niger* (van Hartingsveldt *et al.* (1993) Gene 127, 87-94; Van Gorcom *et al.* (1991) EP420358). On average, the levels obtained with the multicopy cosmid vectors were higher than those obtained with the single copy vectors. Phytase levels obtained with the glucoamylase-phytase fusion vectors resulted in high levels of both glucoamylase and phytase. Controlled batch and fedbatch fermentations from a selected number of phytase producing *A. sojae* transformants revealed a further increased level of phytase.

Glucoamylase production

An example of an efficiently produced fungal protein is provided by the expression of the *A. niger glaA* gene. Vector pGLA6S (Figure 15) is derived from pGLA6 (Punt et al. (1991) J. Biotech. 17, 19-334) by introducing a 5 kb *Eco*RI fragment carrying the *A. nidulans amdS* gene as selection marker into the unique *Eco*RI site of pGLA6. Vector pGLA6S (Figure 15) carrying the *amdS* selection marker and the glucoamlyase gene under control of the *A. nidulans gpdA* promoter was introduced into *A. sojae* ATCC11906pyrG using cotransformation with vector pAB4.1. Starch plate-assays demonstrated the production of increased levels of amylolytic activity in these transformants. From the resulting transformants those showing proficient growth on acrylamide medium were analysed for glucoamylase production. On a Coomassie Briljant Blue-stained SDS PAGE gel from the culture supernatant of some of these transformants a single dominant protein band corresponding to glucoamylase was observed . Western analysis using a monoclonal antibody against glucoamylase (Verdoes *et al.* (1993) Transgenic Research 2, 84-92) was used to confirm the identity of this protein band.

Interleukin-6 production

Production of interleukin-6, which is an example of a highly sensitive protein for proteolytic degradation, was shown to be virtually impossible in *A. niger* without the use of the gla-fusion strategy and protease deficient strains. Even with all these improvements the yields of IL-6 were only a few mg per litre culture fluid. Introduction of the IL-6 vector pAN56-4 (Broekhuijsen *et al.* (1993) J. Biotech. 31, 134-145) into *A. sojae* by cotransformation with the *pyrG* or *amdS* marker resulted in transformants expressing the IL-6-fusion gene present in this vector. From the resulting transformants a few were selected for further analysis. Shake flask fermentation experiments were carried out with these transformants. SDS-PAGE and Western analysis of culture supernatants of several of these strains surprisingly showed levels of correctly processed IL-6 which were about 5-10 fold higher than the levels obtained in the best reported cases in *A. niger*. The use of the various types of protease deficient and fermentation-optimized mutants from *A. sojae* further increased the level of IL-6 production to be obtained from controlled fermentations (Broekhuijsen *et al.* (1993) J. Biotech. 31, 134-145).

Green fluorescent protein (GFP)

Another type of acid labile protein we have attempted to produce in *A.sojae* is GFP from the jelly fish *Aequoria victoria*. This protein is not only proteolytically sensitive but furthermore it loses its activity at acid pH. Vectors carrying GFP or GLA-GFP fusion genes (driven by the *A. nidulans gpdA* promoter) were introduced into *A. sojae* by cotransformation using either the pyrG or amdS selection marker. Expression resulted in brightly fluorescent *A. sojae* transformants for both vector types. Based on the observed fluorescence and the subsequent analysis of culture supernatants from selected, shakeflask-cultured transformants using SDS-PAGE and Western analysis it was ascertained that the yields of intact cytoplasmic GFP and secreted GLA-GFP are much higher than those obtained in *A. niger* protease deficient hosts (Siedenberg *et al.* Biotechn. Prog. (1999) 15, 43-50; Gordon *et al.*, Microbiology (2000) 146, 415-426). In contrast to the situation in *A. niger* culture supernatants also the secreted GFP showed significant fluorescence.

DESCRIPTION OF THE FIGURES

Figure 1: This figure provides a comparison of amino acid sequences of KEX2-like processing proteases from *X. laevis* (XENPC2 and XNFURIN), *S. cerevisiae* (SCKEX2),

*K. lactis* (KLKEX1), *C. albicans* (CAKEX2), *S. pombe* (SPKRP) and *Y. lipolytica* (YLKEX2). The primers, which encode for the amino acid sequences with the highest overall identity (indicated with lightblue boxes), are indicated: MBL793, MBL1208, MBL 794, MBL1158, PE6, PCL1, PCL2(rev), PE6, PCL3, MBL789, PCL4 and MBL1219. Regions of overall identity (4 out of 7 entries) are indicated with purple boxes. Gaps are indicated with . ; no sequence data are indicated with ~ ; asteriks indicate the stop codon of the protein.

Figure 2: This consists of 2a, b and c

Figure 2a provides the background growth of the *A. sojae* strain described in patent WO97/04108 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 2b provides the background growth of the *A. sojae* strain ATCC11906 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 2c provides the background growth of the *A. sojae* strain ATCC20387 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 3 (a and b): This figure provides a comparison of *A. sojae* ATCC11906 and *A. oryzae amdS* sequences from both ends. A and B indicate the two ends. The cloned 1.6 kb *A. sojae* sequence was used. Underlined bold bases differ between species/strains. Intron I sequences are indicated in small letters.

Figure 4 (a and b): This figure illustrates the construction of a pyrG disruption vector via pAO4-13 and pAO4-13deltaCla.

Figure 5: This figure illustrates the construction of pAB4-1rep going from pAB4-1 via

isolation of *Xho*I fragment and *Hind*III fragment followed by cloning into pMTL24.

Figure 6: The construction of the *alpA* gene replacement vector is disclosed in this figure. A 4.4 kb *Eco*RI-*Stu*I fragment from pAS1-1 with the ATCC11906 genomic fragment, the 2.6 kb *Sma*I-*Nco*I fragment from pAB4-1rep and the 4.4 kb *Nco*I-*Eco*RI fragment from pAS1-2A are ligated in a 3 way ligation thus providing pAS1-deltaalp.

Figure 7: This figure provides the restriction map of the DNA fragment carrying the *A. niger pclA* gene.

Figure 8: This figure provides the structure (functional organisation) of the *A. niger pclA* encoded protein. It shows pre, pro, activity and P domains from left to right. The light coloured triangles indicate KR sites. The dark coloured triangles indicate glycosylation sites. The vertically striped light box is an S/P/T rich region. The dark weavepatterned box at the right end is a D/E rich region.

Figure 9: This figure illustrates growth phenotype of an *A. niger pclA* mutant strain.

Figure 10: This figure provides a DNA sequence comparison between the *A. sojae* and *A. niger pclA* genes. A vertical bar indicates identity; : indicates 5; · indicates 1. 72.139% similarity and 72.073% identity were found.

Figure 11: The construction of the *pclA* gene replacement vector is disclosed in this figure. A 7.6 kb *Cla*I fragment, which is a ATCC11906 genomic fragment, was cloned into pMTL23p. In this construct the 2.6 kb *Sma*I fragment from pAB4-1rep was cloned into the *Eco*RV-site, thus providing pAS2-delta pcl.

Figure 12: This figure shows the amino acid sequence comparison of the various PclA homolous from *S. cerevisiae* (Sckex2), *K. lactis* (Klkex1), *A. sojae* (Aspcla), *A. niger* (A. niger), *P. chrysogenum* (Penpcl1), *A. bisporus* (Agarmbl129), *T. reesei* (Trichpcl1), *R. oryzae* (Rhizpcl1), *F. venenatum* (Fuspcl1), *S. pombe* (Spkrp), *C. albicans* (Cakex2) and *Y. lipolytica* (Ylkex2). Regions of overall identity (8 out of 12 entries) are indicated with yellow boxes. Gaps are indicated with .. ; no sequence data are indicated with ~ .

Figure 13: Sequence data are provided in figure 13a for the *A. oryzae alpA* promoter sequences (Q11755). The primer position for PCR cloning is indicated. In figure 13b the sequence data are provided for the *A. oryzae amyA* promoter sequences also including primer positions (A02532). Figure 13c provides the ATCC42149 *A. oryzae* derived *gpdA* promoter sequences (EP0.436.858 a1) also including primer positions.

Figure 14: This figure provides a comparison between various *gpdA* promoter sequences of *Aspergillus*: From top to bottom, *A sojae* ATCC11906, *A. oryzae*, *A. niger* and *A. nidulans*. Asterisks indicate the putative intron present in the 5' untranslated region of the promoters. Arrowhats indicate the CT rich regions. Bold underlined letters indicate the differences between the *A. oryzae* and *A. sojae* sequences.

Figure 15: This figure shows a map of the vector pGLA6S of 12700bp.

SEQUENCE LISTING

**SEQ ID No. 1**

MBL1784: 5'-CGGAATTCGAGCGCAACTACAAGATCAA-3'

**SEQ ID No. 2**

MBL1785: 5'-CGGAATTCAGCCCAGTTGAAGCCGTC-3'

**SEQ ID No. 3**

The sequence of the *Aspergillus niger* gene encoding proprotein convertase
The start codon and the stop codon are indicated with bold underlined letters
The intron is indicated with underlined small letters

```
   1  CCATGGCAAG CCTCCTACTT GGCCTGATTA CATCGTCCTG AGAGAGAGAG

  51  TTCACCAAAA CTCTCCCCCA AACGATGCGT CTTACAGGTG GTGTCGCTGC
```

```
101  GGCTCTGGGC CTCTGCGCTG CTGCCTCCGC TTCTCTCCAT CCCCATCGTT

151  CCTACGAGAC CCATGATTAC TTCGCTCTAC ACCTTGATGA ATCCACCTCG

201  CCGGCCGACG TCGCCCAACG ACTAGGTGCT CGCCACGAAG GCCCCGTCGG

251  AGAATTACCC TCACATCATA CCTTCTCGAT ACCCCGTGAA AACAGTGACG

301  ATGTCCATGC GCTGCTGGAT CAATTGCGCG ATCGTCGGAG GTTACGCCGC

351  CGCTCCGGAG ATGACGCCGC TGTCCTTCCC TCCTTGGTCG GGCGAGACGA

401  AGGTCTAGGT GGCATTCTTT GGTCCGAGAA GCTGGCTCCC CAGAGAAAGC

451  TCCATAAAAG AGTGCCGCCG ACAGGATATG CTGCCAGATC GCCCGTCAAC

501  ACTCAGAATG ACCCCCAAGC GCTTGCGGCG CAGAAACGCA TTGCCTCGGA

551  ATTGGGCATC GCGGACCCCA TCTTCGGCGA ACAATGGCAT TTGTATAATA

601  CTGTTCAGTT GGGCCATGAT CTTAACGTGA CGGGTATCTG GCTGGAGGGC

651  GTTACAGGGC AGGGTGTCAC GACGGCCATT GTCGATGACG GTTTGGACAT

701  GTACAGCAAC GATCTTAGGC CGAACTATTT TGCGGCGGGT TCTTATGACT

751  ATAACGACAA AGTACCAGAG CCGAGGCCGC GCTTGAGCGA TGACCGCCAC

801  GGTACTAGAT GCGCGGGTGA AATCGGTGCG GCGAAGAACG ACGTGTGCGG

851  GGTTGGTGTT GCGTATGATA GTCGCATCGC TGGTATTCGG ATTCTCTCCG

901  CACCCATCGA TGACACTGAT GAGGCTGCGG CTATTAACTA CGCCTATCAG

951  GAGAACGATA TCTACTCGTG TTCCTGGGGT CCCTATGATG ATGGCGCCAC

1001 AATGGAAGCC CCGGGCACTC TGATCAAGCG GGCCATGGTC AATGGTATCC

1051 AAAATGGTCG AGGTGGAAAA GGCTCGGTTT TTGTATTTGC GGCTGGTAAC

1101 GGTGCCATTC ATGACGATAA CTGTAACTTT GACGGTTACA CCAACAGTAT

1151 CTACAGCATC ACGGTGGGTG CCATTGATCG GGAGGGTAAC CATCCTCCGT
```

1201 ATTCGGAATC CTGCTCGGCG CAACTGGTGG TTGCCTACAG CAGCGGCGCC

1251 AGTGATGCAA TTCATACCAC GGACGTCGGC ACAGACAAGT GCTCGACTAC

1301 CCATGGTGGA ACTTCGGCGG CCGGCCCGCT CGCTGCGGGA ACCGTGGCGC

1351 TGGCCCTCAG TGTGCGGCCG GAACTCACCT GGCGTGACGT TCAGTATTTG

1401 ATGATTGAGG CGGCAGTGCC TGTTCATGAA GATGATGGAA GCTGGCAGGA

1451 CACTAAGAAC GGGAAGAAGT TCAGCCATGA CTGGGGATAT GGTAAGGTCG

1501 ACACATATAC GCTGGTGAAA CGGGCAGAGA CCTGGGATCT GGTGAAGCCT

1551 CAAGCCTGGC TCCATTCCCC CTGGCAGCGG GTTGAGCATG AGATCCCACA

1601 GGGCGAGCAG GGCTTGGCTA GTTCGTACGA GGTGACGGAG GATATGTTGA

1651 AGGGAGCCAA CCTGGAACGG CTGGAGCATG TCACGGTCAC CATGAATGTT

1701 AACCACACCC GCCGAGGCGA TCTCAGCGTG GAGTTACGGA GCCCTGATGG

1751 TCGGGTCAGT CACCTCAGTA CGCCCCGGCG GCCAGATAAT CAAGAGGTGG

1801 GCTATGTTGA TTGGACCTTC ATGAGCGTTG CTCACTGgta agtaaaaact

1851 ttttctcggt tgtcggttct tctgctaata catatctagG GGCGAGTCCG

1901 GGATTGGCAA ATGGACTGTG ATTGTCAAGG ACACCAATGT CAACGAGCAT

1951 ACTGGGCAAT TCATCGATTG GCGACTCAAC TTGTGGGGCG AGGCGATTGA

2001 CGGAGCCGAG CAGCCTCTCC ACCCCATGCC TACTGAACAC GATGACGACC

2051 ACAGCTATGA GGAAGGAAAC GTGGCTACCA CGAGCATCAG CGCCGTTCCC

2101 ACGAAAACCG AGCTGCCTGA CAAGCCCACT GGTGCGTTG ATCGCCCGGT

2151 GAACGTTAAG CCTACAACAT CCGCGATGCC GACCGGTAGT CTTACAGAGC

2201 CCATCGATGA TGAAGAACTC CAGAAGACCC CTAGTACAGA GGCAAGCTCA

2251 ACACCAAGTC CTTCTCCGAC CACCGCGTCA GATAGTATCC TGCCTTCCTT

2301 CTTCCCCACG TTCGGTGCGT CGAAGCGGAC CGAAGTTTGG ATCTACGCTG

2351 CGATCGGCTC CATCATTGTG TTCTGCATTG GCCTGGGCGT CTACTTCCAT

2401 GTGCAGCGCC GCAAACGTAT TCGCGACGAC AGCCGGGATG ACTACGATTT

2451 CGAGATGATC GAGGACGAGG ATGAGCTACA GGCAATGAAC GGACGGTCGA

2501 ACCGTTCACG TCGCCGGGGT GGCGAGCTGT ACAATGCTTT TGCGGGCGAG

2551 AGCGATGAGG AACCATTATT CAGTGATGAG GATGATGAAC CGTATCGGGA

2601 TCGGGGGATC AGCGGCGAAC AAGAACGGGA GGGCGCAGAT GGAGAGCATT

2651 CTCGGAGATG AAAGTGCAGT AGATGAGGGT TGACTTTATT TCGGACAGTG

2701 TTTCTAACTT GTTGGATGAC CTGCGTTGAA CAATATTTCT GCTGTGTATG

2751 CTGCATAGAG AAGCGTGTAT ATACCATGTA TGTGTGCATC ATCGTGATCG

2801 GGTTTATCAT TCTTCATCTG CCATGGTTTG TGATCTCCGG AATAGTACCA

2851 AAGGAACACT AAATTAAGGG TCTTGGCGAT GACGCTTCCC GTCGCTGCTT

2901 TTGACTTCCT CCGCATCTCG TCTCTCCTGC TGTTGACCGC GCGCCAACCA

2951 ACCTCCATCT CCTCACTCCT CCCACCTTAA TCTTGCTGTG CTGCTTCTAG

3001 AACCCCCCAG TTTAATTTAA AAACCGGCTT TTCCTAGCTC CACGTATTGT

3051 ACCTCGCACT GATCCCCATC TCCGCCCACT CCAACGCTAC CGACCCAGGC

3101 TTCTCTGGCG GCTCCAGGCG GCAGGCAATC AAACCAACCC CTCGATGGAT

3151 CAGCACGACG ACTTCGACAG SGTCTCGTGG AGGCATGACC CGGACAGCGA

3201 TCTCTCGCGA CCCACGAACT CCGGAACAGA CACAGAGGAA CAGGCGCCAT

3251 ACACTCACGA TGTCAATGGC AAACGGAGGA TGAGCAACCG CTCAAGAAAG

3301 CCCTCAGGCT GGACCACTGG CGGATGCCGT CGACCTGGCG GGCATCGCGA

```
3351  CGGCGTACTA GAGTGTCGGG TAGATTCACC GTTGAAGGAG AATATGGACG

3401  AAAGACGCTT ATATCTCCTA TTTGGTACAC TACTAGGTGG GTATCTTACC

3451  TCAGTGATCT CAGATGGA
```

**SEQ ID No. 4**

The partial sequence in the coding region of the *Aspergillus sojae* gene encoding proprotein convertase

```
   1  CGCGGATCCA TGGAACACGA TGTGCGGGTG AAATTGGAGC AGCTAGGAAT

  51  GATGTCTGTG GAGTAGGTGT TGCATACGAC AGCCAAGTTG CCGGAATTCG

 101  GATTTTGTCC GCACCCATTG ACGACGCAGA TGAGGCTGCT GCCATCAACT

 151  ATGGCTTCCA GCGCAATGAT ATATATTCAT GCTCCTGGGG CCCTCCGGAT

 201  GATGGCGCCA CGATGGAGGC GCCAGGGATT CTTATCAAAC GAGCTATGGT

 251  CAACGGTATC CAAAATGGCC GAGGAGGTAA AGGTTCTATC TTCGTCTTTG

 301  CAGCTGGAAA TGGTGCAGGG TACGATGACA ACTGCAATTT CGACGGTTAT

 351  ACAAACAGCA TTTACAGCAT CACCGTCGGC GCTATTGATC GAGAGGGCAA

 401  ACATCCCAGC TACTCGGAAT CATGCTCTGC CCAGTTGGTT GTCGCTTATA

 451  GCAGTGGCTC GAGTGACGCG ATTCATACCA CCGACGTTGG AACTGATAAA

 501  TGTTATTCAC TNTCACGGGC GGAACTTCTG CAACTGGACC GCTAGCTGCG

 551  GGTACTATTG CCCTCGCTCT TAGTGCCCGA CCGGAACTAA CTTGGCGAGA

 601  TGCCCAGTAC CTGATGATAG AGACCGCAGT TCCCGTCCAC GAAGACGACG

 651  GGAGCTGGCA GACTACCAAA ATGGGGAAGA AGTTTAGCCA TGACTGGGGT

 701  TTTGGGAAAG TAGATGCATA TTCACTGGTC CAGCTGGCCA AGACGTGGGA
```

```
751  GCTGGTGAAA CCACAGGCGT GGTTCCACTC ACCGTGGCTG CGGGTGAAGC

801  ATGAAATCCC ACAAGGTGAC CAGGGCCTTG CCAGCTCATA CGAAATTACC

851  AAGGATATGA TGTACCAGGC CAATGTCGAG AAATTGGAAC ATGTCACTGT

901  GACCATGAAT GTAAATCACA CTCGCCGAGG CGATATCAGC GTGGAGTTGC

951  GCAGCCCCGA AGGTATCGTC AGTCATCTGA GTACAGCGCG GCGGTCAGAT

1001 AATGCAAAGG CTGGCTATGA AGATTGGACG TTTATGACTG TGGCTCATTG

1051 GTATGTATTT GCTCCCGTAA TTTAGTTTTC GTGCTCAGTC CTGACATTTA

1101 CATTTAGGGG TGAGTCCGGT GTTGGAAAGT GGACGGTCAT TGTGAAGGAT

1151 ACCAATGTCA ATGATCATGT TGGAGAATTC ATCGACTGGC GGCTCAACCT

1201 CTGGGGACTT TCGATCGACG GCTCCAGCCA GCCCCTTCAT CCTATGCCCG

1251 ATGAGCATGA CGATGACCAC TCGATTGAAG ATGCCATTGT TGTTACCACT

1301 AGTGTTGACC CTATCCCAAC TAAGACTGAA GCCCCACCTG TCCCAACTGA

1351 TCCCGTGGAT CGTCCTGTGA ACGCAAAGCC ATCTGCGCAG CCAACGATGC

1401 CTTCAGAGGC TCCTGCTCAA GAGACATCTG AAGTTCCCAC CCCGACGAAA

1451 CCTAGTTCTA CTGAATCACC TTCTTACCAC CTCCTCTGCG GATAGCTTTT

1501 TGCCATCCTT CTTCCCCACG TTCGGTGCGT CGTGAGGATC CAAGCTTGGG

1551 TACGT
```

**SEQ ID No. 5**

The partial sequence in the coding region of the *Trichoderma reesei* QM9414 gene encoding proprotein convertase

```
1  GCTGTCCGCA CTGATGCGTG CGGCCTTGGC GTTGCCTACG ACTCCAAGAT
```

```
  51  TGCTGGCATC CGCATCCTTA GTAGTGCCAT CAGCGATGCG GACGAGGCCG

 101  AGGCCATGAT TTACAAGTTC CAGGACAACC AGATCTACTC GTGCTCCTGG

 151  GGGCCTCCCG ACGATGGGAG GTCCATGGAA GCCCCCGACG TCCTGATTCG

 201  ACGAGCCATG CTCAAGGGCG TCCAGGAGGG ACGAGGAGGC CTCNGCTCCA

 251  TCTACGNCTT TGCTAGTGGT AACGGTGCCG CCAGTGGCGA TAACTGCAAC

 301  TNCGACGGAT ACNCAAACA
```

**SEQ ID No. 6**

The partial sequence in the coding region of the *Fusarium venenatum* ATCC20334 gene encoding proprotein convertase

```
   1  GGTTTNNCCG TTGGTGTTGC TACGACTCCA AGTCGCCGGA ATCCGTATTC

  51  TCAGCAAACT GATCAGCGAC GCCGACGAAG CAGAAGCGCT TATGTACAAG

 101  TACCATGACA ACCATATTTA CTCTTGCTCA TGGGGTCCTT CCGATGATGG

 151  CCAGACTATG GAGGCACCCG ATGTTGTCAT TCGACGAGCA ATGCTTAAGG

 201  CGATTCAGGA GGGACGTAAT GGTCTTGGCT CTGTCTACGT CTTTGCCAGT

 251  GGAAACGGTG CAGGCCAAGG AGATAACTGC AACTNCGACG GATCCACCAA

 301  ACA
```

SEQ ID No. 7

The partial sequence in the coding region of the *Penicillium chrysogenum* P2 gene encoding for proprotein convertase

```
  1  GTGGGTGTTG CCTATGACAG CAAGGTGTCA GGTATCCGGA TTCTGTCCAA

 51  GGCGATTGAC GACGTCGACG AAGCAGCTGC CATCAACTTT GCCTTCCAAG

101  ATAACGATAT ATACTCCTGC TCGTGGGGTC CTCCTGATGA TGGTGCGACC

151  ATGGATGCGC CGGGCTTGTT GATCAAGCGG GCGATGGTCA ATGGTGTGCA

201  NGAGGGACGA GGTGGAAAGG GTTCGATCTT CGTGTTNGCC GCAGGCAACG

251  GTGCTCTTTT TGGCGACAAC TGCAACTTCG ACGGATACAA CAAACA
```

**SEQ ID No. 8**

The partial sequence in the coding region of the *Rhizopus oryzae* ATCC200076 gene encoding proprotein convertase

```
  1  ACTNGGGGCA TTGGTGAAAT NTTGCTTGTG GNTTGGTGTT GCTTACGACG

 51  CAAAAATATC TGGTATACGT ATATTATCAG GTGAAATCAC AGAGGCAGAC

101  GAGGCTGCTG CTTTGAATTA CAAATATCAA GAAAATCAAA TCTACTCCTG

151  CTCNTGGGGC CCA
```

**SEQ ID No. 9**

The partial sequence in the coding region of the *Agaricus bisporus* HORST gene encoding proprotein convertase

```
  1  ATGTGGTCTT GGTCTCGCCT ACGAATCCAA GGTCGCTGGT GTTCGCATAT

 51  TGTCTGGTCC CATAACGGAC GTCGATGAAG CGACTGCGCT CAACTATGGT

101  TTCCAAAATG TATCTATCTT CAGCTGTAGT TGGGGCCCAC CTGACAATGG

151  TATGTCCATG GAAGGCCCAG GATACCTCAT CAAAAAAGCT GTCGTCAACG
```

201 GCATTAACCA GGGACGTGGC GGGAAGGGCT CCATTTTCGT CTTCGCCAGT

251 GGCAACGGCG CTGCTTCGGA TGACCAATGC AACTACGACG GATACACAAA

301 CA

**SEQ ID No. 10**

coding strand *Bam*HI-site is underlined

PE4 5'- CG CGGATC CA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GG -3'

degenerated 2048 times

**SEQ ID No. 11**

coding strand

PCL1 5'-CA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GGX GA-3'

degenerated 8192 times

**SEQ ID No. 12**

coding strand

PCL2 5'-AT(C/T/A) TA(T/C) TCX TG(T/C) TCX TGG GGX CC-3'

degenerated 768 times

**SEQ ID No. 13**

non coding strand *Bam*HI-site is underlined

PE6 5'- CGC GGA TCC XCC (A/G)TT XCC X(C/G)(A/G) XGC (G/A/C)(C/A)A XAC -3'

degenerated 49152 times

**SEQ ID No. 14**

non coding strand

PCL2rev 5'-GG XCC CCA XGA (A/G)CA XGA (A/G)TA (A/T/G)AT-3'

degenerated 768 times

**SEQ ID No. 15**

non coding strand

PCL3 5'-(A/G)TT XGT (A/G)TA XCC (A/G)TC (A/G)(A/T)A (A/G)TT-3'

degenerated 1024 times

**SEQ ID No. 16**

non coding strand

PCL4 5'-GC XGC XGA XGT XCC XCC (A/G)TG-3'

degenerated 2048 times

**SEQ ID No. 17**

The sequence of pAB4-1rep

.....59-499 bp.....: 0.4 kb *Hind*III fragment of pAB4-1

1-58 bp.....500-513 bp.....2873-2930 bp : polylinker sequence of pMTL24 (indicated with underlined small letters)

.....514-2872 bp.....: 2.3 kb *Xho*I fragment of pAB4-1

```
   1  ggccagtgaa ttcgagctcg gtacccgggg atcctctaga gtcgacctgc
  51  aggcatgcAA GCTTGGTCAG CAGTACCAGA CGCCCGGATC GGCTATCGGC
 101  CGGGGTGCTG ACTTCATTAT CGCGGGTCGC GGTATCTACG CCGCGCCGGA
 151  TCCGGTGCAG GCTGCGCAAC AGTATCAGAA GGAGGGGTGG GAAGCCTACC
 201  TGGCCCGTGT CGGCGGAAAC TAATACTATA AAAGGAGGAT CGAAGTTCTG
 251  ATGGTTATGA ATGATATAGA AATGCAACTT GCCGCAACGG ATACGGAAGC
 301  GGAAACGGAC CAATGTCGAG CACGGGTAGT CAGACTGCGG CATCGGATGT
 351  CCAAACGGTA TTGATCCTGC AGGCTACTAT GGTGTGGCAC AAGGATCAAT
 401  GCGGTACGAC GATTTGATGC AGATAAGCAG GCTGCGAAGT AGTAACTCTT
 451  GCGTAGAGAA AATGGCGACG GGTGGGCTGA TAAGGGCGGT GATAAGCTTg
```

```
 501  catgcctgca ggcCTCGAGC TAACATACAT TCCGAACCGT GCAGCCCAAG

 551  GCCGAGCAGT TCAACTGCGC TCAGCGCGCT CATGCCAACT TCCTTGAGAA

 601  CTCCAGCCAA ACTATGCTCT TCCTCCTGGT AGCTGGACTG AAGTACCCCC

 651  AGTTGGCGAC TGGCCTCGGA AGCATCTGGG TCCTCGGTCG CTCACTGTTC

 701  CTTTACGGAT ATGTGTACTC CGGCAAGCCG CGGGGTCGCG GTCGTTTGTA

 751  CGGCAGCTTC TACTTGCTTG CACAGGGAGC TCTCTGGGGC NTGACGTCTT

 801  TTGGAGTTGC GAGGGAGTTG ATTTCCTACT TCTAAGTTTG GACTTGAATC

 851  CGTGGTGTGA TTGAGGTGAT TGGCGATGTT TGGCTATACC AGCTATATGT

 901  AATAATCTCT ACTGTATACT ACTATTCAAC GCATTTTACT ATGCGTGCTG

 951  CTAGGGTCGG CAATGACAAT GGCAATCTGA CTGACGTGGT CTATTTCTCC

1001  ATGTGCAGCA GGGAATACGA GCTCCAATGG ACCTCGGGAG TGGCACAGTC

1051  AATGGCAAGG AAACTCCGCC TTTGCAGGTG TGGCTGAACC CCACGGGTCG

1101  GAGGCGGAGC AATCCACCCC CGATGTGGCT GGTGCGTGGA GGGGCTCGCG

1151  ATGATTTTAC TGAGCTTGCT TTTCTTGTCG ACATTGAACA TTGTCCTTGG

1201  TCTTCCTTCA GATTTAAGGG TCAGTCACTG CTACATTTCT CAGTAGTATC

1251  CGCGCACGTC TCTGGATTTA CGAATCAGGG TCCACCAGTC GAAACTTCGA

1301  ACTACTCTCA TTATACAATC CTCTTTCCAT TCCCGCATTA ACCCCTCCAT

1351  CAACACCATG TCCTCCAAGT CGCAATTGAC CTACACTGCC CGTGCCAGCA

1401  AGCATCCCAA TGCTCTGGCG AAGAGGCTGT TCGAGATTGC CGAGGCCAAG

1451  AAGACCAATG TGACTGTCTC GGCTGACGTT ACCACCACTA AGGAGCTACT

1501  AGATCTTGCT GACCGTAGGC CGACCCGCTA CTCTGCCTGA TTATGCTGCA

1551  TGCAAACTTA TTAACGGTGA TACCGGACTG CAGGTCTCGG TCCCTACATT
```

1601 GCCGTGATCA AAACCCACAT CGATATCCTC TCTGATTTCA GCAACGAGAC

1651 CATTGAGGGA CTTAAGGCTC TCGCGCAGAA GCACAACTTT CTCATCTTCG

1701 AGGACCGCAA GtTCATTGAC ATCGGCAACA CGGTCCAGAA GCAATACCAC

1751 GGCGGTACCC TCCGTATCTC GGAATGGGCC CACATCATCA ACTGCAGCAT

1801 TCTCCCTGGT GAGGGTATCG TCGAGGCTCT CGCTCAGACG GCGTCTGCAC

1851 CGGACTTCGC CTACGGCCCC GAACGCGGTC TGTTGATCTT GGCAGAGATG

1901 ACCTCTAAGG GCTCCTTGGC TACCGGCCAG TACACTACTT CCTCGGTCGA

1951 TTATGCCCGG AAATACAAGA ACTTCGTTAT GGGATTCGTG TCGACGCGCG

2001 CGTTGGGTGA GGTGCAGTCG GAAGTCAGCT CTCCTTCGGA TGAGGAGGAC

2051 TTTGTGGTCT TCACGACTGG TGTGAACATT TCTTCCAAGG GAGATAAGCT

2101 TGGTCAGCAG TACCAGACGC CCGGATCGGC TATCGGCCGG GGTGCTGACT

2151 TCATTATCGC GGGTCGCGGT ATCTACGCCG CGCCGGATCC GGTGCAGGCT

2201 GCGCAACAGT ATCAGAAGGA GGGGTGGGAA GCCTACCTGG CCCGTGTCGG

2251 CGGAAACTAA TACTATAAAA GGAGGATCGA AGTTCTGATG GTTATGAATG

2301 ATATAGAAAT GCAACTTGCC GCAACGGATA CGGAAGCGGA AACGGACCAA

2351 TGTCGAGCAC GGGTAGTCAG ACTGCGGCAT CGGATGTCCA AACGGTATTG

2401 ATCCTGCAGG CTACTATGGT GTGGCACAAG GATCAATGCG GTACGACGAT

2451 TTGATGCAGA TAAGCAGGCT GCGAAGTAGT AACTCTTGCG TAGAGAAAAT

2501 GGCGACGGGT GGGCTGATAA GGGCGGTGAT AAGCTTAATT GTCATCGCAG

2551 ATAAGCACTG CTGTCTTGCA TCCAAGTCAG CGTCAGCAGA AATACGGGAC

2601 TTCCGAAAGT ATATGGCAAA ATTAAAGAAC TTGACTCTCC AGCAATGTTT

2651 TGCCCTGACC GTCGCTAAAA CGTTACTACC CCTATACCCG TCTGTTTGTC

2701 CCAGCCCGAG GCATTAGGTC TGACTGACAG CACGGCGCCA TGCGGGCTTG

2751 GGACGCCATG TCCGTCGCGT GATAAGGGTT GATCCATGCA GCTACTATCC

2801 TTCCATCGTT CCATTCCCAT CCTTGTCCTA TCTCCATCCT TGAAACTTTA

2851 CTAGTTTAGT TGGATGCTCG AGatctccat ggacgcgtga cgtcgactct

2901 gaggatcccc gggtaccgag ctcgaattcg

**SEQ ID No. 18**

MBL 789 *Eco*RI is underlined

5'- GGAA TTC (A/G)GA ATA (T/A)GG AGG ATG TAG -3'

degenerated 4 times

**SEQ ID No. 19**

MBL 793 *Bam*HI is underlined

5'- CGGATCCG CAG TGG CAC TTG (G/A)TC AAT CCA A -3'

degenerated 2 times

**SEQ ID No. 20**

MBL 794 *Eco*RI is underlined

5'- GGA ATT CTT AAA A(T/G)C CCA AGA ACC TTC A -3'

degenerated 2 times

**SEQ ID No. 21**

MBL 1158 *Eco*RI is underlined

5'- G GAA TTC (T/C)TC (T/G)CC (T/G)GC (A/G)CA (C/G)C(T/G) (C/G)GT (T/G)CC (A/G)TG -3'

degenerated 512 times

**SEQ ID No. 22**

MBL 1208 *Cla*I is underlined

5'- CGG ATC GA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GG -3'
degenerated 2048 times

**SEQ ID No. 23**

MBL 1219 *Bam*HI is underlined

5'- CGG ATC (C/T)TG XA(G/T/C) (A/G)TC XC(T/G) CCA XGT (C/A/G)AG -3'
degenerated 4608 times

**SEQ ID No. 24**

Restriction sites are bold
Primers are underlined

*Bam*HI PE4 primer
GGATCCATGG CACGAGATGT GCAGGTGAAA TCGGTGCGGC GAAAGAAAAC AACGTGTGCG 60

GGGTTGGTGT TGCGTATGAT AGTCGCATCG CTGGTATTCG GATTCTCTCC ACACCCATCG 120

*Eco*RV
ATGACACTGA TGAGGCTGCG GCTATTAACT ACGCCTATCA GGAGAAC**GAT ATC**TACTCGT 180

GTTCCTGGGG TCCCTATGAT GATGGCGCCA CAATGGAAGC CCCGGGCACT CTGATCAAGC 240

GGGCCATGGT CAATGGTATC CAAAATGGTC GAGGTGGAAA AGGCTCGGTT TTTGTCTGCG 300

PE6 primer
CCCCCGGAAA TGGT**GGATCC**
320
*Bam*HI

**SEQ ID No. 25**

*Aspergillus niger* PclA protein sequence

1 Met Arg Leu Thr Gly Gly Val Ala Ala Ala Leu Gly Leu Cys Ala

16 Ala Ala Ser Ala Ser Leu His Pro His Arg Ser Tyr Glu Thr His

31 Asp Tyr Phe Ala Leu His Leu Asp Glu Ser Thr Ser Pro Ala Asp

46 Val Ala Gln Arg Leu Gly Ala Arg His Glu Gly Pro Val Gly Glu

61 Leu Pro Ser His His Thr Phe Ser Ile Pro Arg Glu Asn Ser Asp

76 Asp Val His Ala Leu Leu Asp Gln Leu Arg Asp Arg Arg Arg Leu

91 Arg Arg Arg Ser Gly Asp Asp Ala Ala Val Leu Pro Ser Leu Val

106 Gly Arg Asp Glu Gly Leu Gly Gly Ile Leu Trp Ser Glu Lys Leu

121 Ala Pro Gln Arg Lys Leu His Lys Arg Val Pro Pro Thr Gly Tyr

136 Ala Ala Arg Ser Pro Val Asn Thr Gln Asn Asp Pro Gln Ala Leu

151 Ala Ala Gln Lys Arg Ile Ala Ser Glu Leu Gly Ile Ala Asp Pro

166 Ile Phe Gly Glu Gln Trp His Leu Tyr Asn Thr Val Gln Leu Gly

181 His Asp Leu Asn Val Thr Gly Ile Trp Leu Glu Gly Val Thr Gly

196 Gln Gly Val Thr Thr Ala Ile Val Asp Asp Gly Leu Asp Met

Tyr

211 Ser Asn Asp Leu Arg Pro Asn Tyr Phe Ala Ala Gly Ser Tyr Asp

226 Tyr Asn Asp Lys Val Pro Glu Pro Arg Pro Arg Leu Ser Asp Asp

241 Arg His Gly Thr Arg Cys Ala Gly Glu Ile Gly Ala Ala Lys Asn

256 Asp Val Cys Gly Val Gly Val Ala Tyr Asp Ser Arg Ile Ala Gly

271 Ile Arg Ile Leu Ser Ala Pro Ile Asp Asp Thr Asp Glu Ala Ala

286 Ala Ile Asn Tyr Ala Tyr Gln Glu Asn Asp Ile Tyr Ser Cys Ser

301 Trp Gly Pro Tyr Asp Asp Gly Ala Thr Met Glu Ala Pro Gly Thr

316 Leu Ile Lys Arg Ala Met Val Asn Gly Ile Gln Asn Gly Arg Gly

331 Gly Lys Gly Ser Val Phe Val Phe Ala Ala Gly Asn Gly Ala Ile

346 His Asp Asp Asn Cys Asn Phe Asp Gly Tyr Thr Asn Ser Ile Tyr

361 Ser Ile Thr Val Gly Ala Ile Asp Arg Glu Gly Asn His Pro Pro

376 Tyr Ser Glu Ser Cys Ser Ala Gln Leu Val Val Ala Tyr Ser Ser

391 Gly Ala Ser Asp Ala Ile His Thr Thr Asp Val Gly Thr Asp Lys

406 Cys Ser Thr Thr His Gly Gly Thr Ser Ala Ala Gly Pro Leu Ala

421 Ala Gly Thr Val Ala Leu Ala Leu Ser Val Arg Pro Glu Leu Thr

436 Trp Arg Asp Val Gln Tyr Leu Met Ile Glu Ala Ala Val Pro Val

451 His Glu Asp Asp Gly Ser Trp Gln Asp Thr Lys Asn Gly Lys Lys

466 Phe Ser His Asp Trp Gly Tyr Gly Lys Val Asp Thr Tyr Thr Leu

481 Val Lys Arg Ala Glu Thr Trp Asp Leu Val Lys Pro Gln Ala Trp

496 Leu His Ser Pro Trp Gln Arg Val Glu His Glu Ile Pro Gln Gly

511 Glu Gln Gly Leu Ala Ser Ser Tyr Glu Val Thr Glu Asp Met Leu

526 Lys Gly Ala Asn Leu Glu Arg Leu Glu His Val Thr Val Thr Met

541 Asn Val Asn His Thr Arg Arg Gly Asp Leu Ser Val Glu Leu Arg

556 Ser Pro Asp Gly Arg Val Ser His Leu Ser Thr Pro Arg Arg Pro

571 Asp Asn Gln Glu Val Gly Tyr Val Asp Trp Thr Phe Met Ser Val

586 Ala His Trp Gly Glu Ser Gly Ile Gly Lys Trp Thr Val Ile Val

601 Lys Asp Thr Asn Val Asn Glu His Thr Gly Gln Phe Ile Asp Trp

616 Arg Leu Asn Leu Trp Gly Glu Ala Ile Asp Gly Ala Glu Gln Pro

631 Leu His Pro Met Pro Thr Glu His Asp Asp Asp His Ser Tyr Glu

646 Glu Gly Asn Val Ala Thr Thr Ser Ile Ser Ala Val Pro Thr Lys

661 Thr Glu Leu Pro Asp Lys Pro Thr Gly Gly Val Asp Arg Pro Val

676 Asn Val Lys Pro Thr Thr Ser Ala Met Pro Thr Gly Ser Leu Thr

691 Glu Pro Ile Asp Asp Glu Glu Leu Gln Lys Thr Pro Ser Thr Glu

706 Ala Ser Ser Thr Pro Ser Pro Ser Pro Thr Thr Ala Ser Asp Ser

721 Ile Leu Pro Ser Phe Phe Pro Thr Phe Gly Ala Ser Lys Arg Thr

736 Glu Val Trp Ile Tyr Ala Ala Ile Gly Ser Ile Ile Val Phe Cys

751 Ile Gly Leu Gly Val Tyr Phe His Val Gln Arg Arg Lys Arg Ile

766 Arg Asp Asp Ser Arg Asp Asp Tyr Asp Phe Glu Met Ile Glu Asp

781 Glu Asp Glu Leu Gln Ala Met Asn Gly Arg Ser Asn Arg Ser Arg

796 Arg Arg Gly Gly Glu Leu Tyr Asn Ala Phe Ala Gly Glu Ser

Asp

811 Glu Glu Pro Leu Phe Ser Asp Glu Asp Asp Glu Pro Tyr Arg Asp

826 Arg Gly Ile Ser Gly Glu Gln Glu Arg Glu Gly Ala Asp Gly Glu

841 His Ser Arg Arg

**SEQ ID Nos. 26 to 31**

PCR-primers for *A.sojae* promoter cloning
Restriction sites are underlined

| | Primer | Sequence (5' - 3') |
|---|---|---|
| SEQ ID No. 26 | Alp-1 | GGAATTCGCGGCCGCGGTTATTCTGCGGAAGC G *EcoRI* *NotI* |
| SEQ ID No. 27 | Alp-2 | GGAATTCCCATGGTGAGAAGATTGTAAAG *EcoRI* *NcoI* |
| SEQ ID No. 28 | Amy-1 | GGAATTCGCGGCCGCAGATCTGCCCTTATAAA TCTCC *EcoRI* *NotI* |
| SEQ ID No. 29 | Amy-2 | GGAATTCCCATGGATGCCTTCTGTGGGG *EcoRI* *NcoI* |
| SEQ ID No. 30 | AOGPDA -1 | GGAATTCGCGGCCGCCTATGAAACCGGAAAG *EcoRI* *NotI* |
| SEQ ID No. 31 | AOGPDA -2 | GGAATTCTAGCCATGGTTTAGATGTG *EcoRI* *NcoI* |

**SEQ ID No. 32**

The sequence of the *Aspergillus sojae gpdA* promoter region

```
   1  AATTGCGGCC GCTATGAAAC CGGAAAGGGC TGCTGAGAGC TGGGGAACGG

  51  CGCAAGCCGG GAAAACAGCT GACAAGGACC CATTTCACTC TGGATCTTGA

 101  GGAGAGCTGT AGCTTTTGCC CCGTCTGTCC ACCCGGTGAC TGGATTAGTG

 151  ACCTGGTCGT TGCGTCAGTC AACATTGCTC TTTTTTTATC TCCCCCTCCC

 201  CCGCCGTCCG ACTTTTCTCC CCTTTTCTAC TCTCTTCGTA TACTCACCAC

 251  TGCAATCATC TTATCCCTTT GTCTTCTTAC TTAAAGTGAG TCGTCTCCCG

 301  CCCATCGTTC CCTTTGAACC TTGTAAATCA GAGCCACTTT CAAGTGTCTA

 351  CCGTTTCCTT TCCACATAGA TTGACTGACA GCTACCCCGC CACACCAGCA

 401  GACACATCTA AACCATGG
```

**SEQ ID No. 33**

The sequence of the *Aspergillus sojae alpA* promoter region

```
   1  GCGGCCGCGG TTATTCTGCG GAAGCGGACC CCCCCCTTCC GCCCAAACAG

  51  GGCGAATGTG CCCAAGTTCT GATACTATCA GAAGACCTCC AGGAGCACAT

 101  GCCTGTTCGC ATAACCCTGG TGTAGCACCA GGAATTGCTT AGCTTAGCTT

 151  CTTCGACTGA GGGGCCAGAA AGTGCTTATC GCAAAGATCC CACTTCTTTG

 201  TGTGATAGCC CCTCCCGCGG CCCTTGATCA AGCCGTTCTC GCTATCCAAT

 251  ATTGAAAGCG TGATATTATA GGTGCACATG GTTATTATCC TTTTTCTTTT

 301  TCTCTTTCTT TGCTTTTCAT GCAACCCCAT ACGTTGCCGA ATTTGGCTAC

 351  ACCTTGGGGC TCATTCTTCG AAGTTTAGAT TCCGACAAGA CCTCACCACC

 401  CAATCAAAAC CCTTGATTCC TGATAAAAGA CGTGGAAAGA AGCGGATATC

 451  GCGTGAGGAT GCCAAGCAAA GGGAATGGGT CACATTGATC TCTGTCGCGT
```

```
 501  TGTTAGGATG ATCTTCACTC CTAAAGGCAT CGCCCGCGGC ACTAGGTCCT

 551  TCCTGTCCAG GATATCGTTT ACTCCTCTCA TTATGGCGAG CTACTTTGTG

 601  AATTAATTGA CTGAGGGATA TACCACCTTC CCTTTGAAGG TACCAAGCCA

 651  CTACCTTGAG CGTTAGTTAC TTTTTCGAGG AAAGCGTCCT ATGCTGGTCT

 701  CCGCCAAACC CTCGACAACT TGCCATAGCC TTGTGTTCTT CATGGTCTAT

 751  CGGAGTACCC GTTCATGACT GAAGCGGGTC AGCGTCCGTG GTGGTCATCA

 801  TCATTCTCAT CTTTCATCAT GCCCGCTGAT TGATAGAGTA ATTTCCGGTG

 851  GAGCACAACG CCGTCCTCTG AGATGCAATG TCACCCTGTA AGTTTCAACT

 901  ACACTCTGTA GTACAGAGCA TCCTTGCCAT TGCATGCTGT GCAAGTGATC

 951  TAAATCCGTA GAATCTGCTC GAGAACGGGG AAATATAGAA CTCCTGAAGG

1001  TTATAAATAC CACATGCATC CCTCGTCCAT CCTCATTTCC ATCATCAAGC

1051  CAGCGGTTTC TATCCTCCGA CTTGAGTCGT TCTCGCGCAT CTTTACAATC

1101  TTCTCACCAT GG
```

**Table 1. Taxonomic scheme of the genus *Aspergillus* (Samson, 1992)**

| GENUS | SUBGENUS | SECTION | SELECTED SPECIES[a] "SUBSPECIESES"[b] | |
|---|---|---|---|---|
| Aspergillus | Circumdati | Wentii | A. wentii (glucosidase) | |
| | | Flavi/Tamarii | A.oryzae (amylase, protease) | |
| | | | A. tamarii[tox] | |
| | | | **A. sojae** (fermented food, protease) | |
| | | | A. parasiticus[tox] | |
| | | | A. flavus[tox] | |
| | | Nigri | A. niger ----------→ | A. pulverulentes |
| | | | (fermented food, various proteins, organic acids) | A. phoenicis |
| | | | | A. awamori |
| | | | | A. foetidus |
| | | | | A. kawachii |
| | | | | A. usamii |
| | | | | A. ficuum |
| | | | A. japonicus----------→ | A. aculeatus |
| | | | (endoglucanase) | (glucosidase, galactanase) |
| | | | A. ellipticus | |
| | | | A. tubingensis ----------→ | "A. niger" |
| | | Circumdati | A. ochraceus[tox] (xulanase) | |
| | | | A. alliaceus[tox] | |
| | | Candidi | A. candidus (lipase, glucosidase) | |
| | | Cremei | A.itaconicus (organic acid) | |
| | | Sparsi | A. sparsus | |
| Aspergillus | Aspergillus | Aspergillus | A. glaucus (fermented food) | |
| | | Restricti | A. restrictus[tox] | |
| | Fumigati | Fumigati | A. fumigatus[tox] | |
| | | Cervini | | |
| | Ornati | | | |
| | Clavati | Clavati | A. giganteus[tox] | |

| | | |
|---|---|---|
| Nidulantes | Nidulantes | A. nidulans[tox] |
| | Versicolores | A. sydowii (lipase) |
| | Usti | |
| | Terrei | A. terreus[tox] (glucansae) |
| | Flavipedes | |

[a] For the species selected for this list either the production of proteins/organic acids/fermented foods (indicated between brackets) and/or a DNA-mediated transformation procedure (indicated by underlining) is described, except for *A. tamarii*, *A. sparsus* and *A. ellipticus*. Species recorded to produce toxins are indicated by [tox]

[b] Based on various methods the listed names may be considered synonymous to the given SPECIES name.

**Table 2. The classificati n of the different ATCC strains**

| Strains | Morphology[1)] | Aflatoxin Production | RAPD[2)] | $PCR_{aflR}$ [3)] | $PCR_{alpA}$ | Classification[4)] |
|---|---|---|---|---|---|---|
| ATCC 9362[5)] | ND | no[7)] | *A. sojae* type I | *A. sojae* | *A. sojae* | *A. sojae* |
| ATCC 11906[6)] | *A. sojae* | no[1,7)] | *A. sojae* type I | ND | *A. sojae* | *A. sojae* |
| ATCC 20235 | *A. oryzae* | no[1)] | *A. sojae* type II | ND | *A. oryzae* | *A. oryzae* |
| ATCC 20245 | *A. sojae* | no[1,7)] | *A. sojae* type I | *A. sojae* | *A. sojae* | *A. sojae* |
| ATCC 20387 | *A. sojae* | no[1)] | ND | ND | *A. sojae* | *A. sojae* |
| ATCC 20388 | *A. sojae* | no[1)] | ND | ND | *A. sojae* | *A. sojae* |
| ATCC 42249 | *A. sojae* | no[1)] | *A. sojae* type II | ND | *A. sojae* | *A. sojae* |
| ATCC 42250 | *A. sojae* | no[1)] | ND | ND | *A. sojae* | *A. sojae* |
| ATCC 42251 | *A. sojae* | no[1)] | ND | *A. sojae* | *A. sojae* | *A. sojae* |
| ATCC 46250 | ND | ND | ND | ND | *A. oryzae* | *A. oryzae* |
| IFO 4177 (CBS 205.89) | *A. oryzae* | No[8)] | ND | ND | *A. oryzae* | *A. oryzae* |

Legend : ND = not determined

[1] REF: **Ushijima S, Hayashi K and Murakami H** (1982) The current taxonomic status of *Aspergillus sojae* used in *Shoyu* fermentation. Agric. Biol. Chem., 46:2365-2367, 1981.

[2] REF: **Yuan GF, Liu CS and Chen CC** (1995) Differentiation of *Aspergillus parasiticus* from *Aspergillus sojae* by Random Amplification of Polymorphic DNA. Appl. Environm. Microbiol., 61:2384-2387.

[3] REF: **Chang PK, Bhatnagar D, Cleveland TE and Bennett JW** (1995) Sequence variability in homologs of the aflatoxin pathway gene *aflR* distinguishes species in *Aspergillus* section *Flavi*. Appl. Environm. Microbiol., 61:40-43.

[4] Conclusion on classification drawn by TNO based on data presented in this table

[5] This strain was deposited at ATCC as *A. oryzae*, but later reclassified as *A. sojae* based on Yuan et al, 1995[2] and Chang et al, 1995[3]

[6] This strain was deposited at ATCC as *A. parasiticus*, but later reclassified as *A. sojae* based on Ushijima et al, 1981[1] and Yuan et al, 1995[2]

[7] REF: ATCC catalogue

[8] REF: **Liu BH, Chu FS** (1998) Appl. Env. Microbiol., 64:3718-3723.

Table 3. Composition of selection media

| Composition | Non-selection medium | Selection medium (WO97/04108) | Acrylamide selection medium | Improved acrylamide selection medium |
|---|---|---|---|---|
| $KH_2PO_4$ | 1.5 g/l | 1.5 g/l | 1.5 g/l | 1.5 g/l |
| KCl | 0.5 g/l | 0.5 g/l | 0.5 g/l | 0.5 g/l |
| $MgSO_4.7H_2O$ | 0.5 g/l | 0.5 g/l | 0.5 g/l | 0.5 g/l |
| $NaNO_3$ | 6 g/l | ---- | ---- | ---- |
| glucose | 10 g/l | 10 g/l | 10 g/l | ---- |
| sorbitol | 1.2 M | ---- | 1.2 M | 1.2 M |
| saccharose | ---- | 1 M | ---- | ---- |
| mineral solution[1)] | 0.1% v/v | 0.1% v/v | 0.1% v/v | 0.1% v/v |
| acetamide | ---- | 10 mM | ---- | ---- |
| acrylamide | ---- | ---- | 10 mM | 10 mM |
| CsCl | ---- | 15 mM | 15 mM | 15 mM |
| agar | 15 g/l | 15 g/l | 15 g/l | 15 g/l |

| | | |
|---|---|---|
| [1)] mineral solution : | $CuSO_4.5H_2O$ | 0.16 g/l |
| | $FeSO_4.7H_2O$ | 0.5 g/l |
| | $ZnSO_4.7H_2O$ | 2.2 g/l |
| | $MnCl_2.4H_2O$ | 0.5 g/l |
| | $CoCl_2.6H_2O$ | 0.17 g/l |
| | $Na_2MoO_4.2H_2O$ | 0.15 g/l |
| | $H_3BO_3$ | 1.1 g/l |
| | EDTA | 5 g/l |

**Table 4. Protease activity in different media**

| STRAINS | Degradation of proteins after incubation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Minimal Medium | | | Complete Medium | | | Minimal Medium + Trusoy | | | Skim milk |
| | PH | BSA | Phytase (*A.terreus*) | pH | BSA | Phytase (*A.terreus*) | pH | BSA | Phytase (*A.terreus*) | |
| ATCC 9362 | 6.75 | - | - | 7.5 | - | - | 6.88 | - | - | - |
| ATCC 11906 | 7.00 | - | - | 8.28 | - | - | 7.51 | - | - | - |
| ATCC 20235 (= *A. oryzae*) | 8.40 | - | + | 8.38 | + | + | 7.70 | + | + | + |
| ATCC 20245 | 8.05 | - | + | 8.18 | - | + | 7.65 | + | + | + |
| ATCC 20387 | 7.45 | - | - | 7.5 | + | + | 7.76 | - | - | - |
| ATCC 20388 | 8.20 | - | + | 7.5 | - | + | 7.74 | - | + | + |
| ATCC 42249 | 8.30 | - | + | 7.5 | - | + | 7.77 | + | + | + |
| ATCC 42250 | 8.30 | - | + | 7.5 | - | + | 7.55 | - | + | + |
| ATCC 42251 | 8.40 | - | + | 7.5 | - | + | 7.50 | - | + | + |
| ATCC 46250 (= *A. oryzae*) | 7.75 | - | + | 8.0 | + | + | 7.35 | + | + | + |
| *A.niger* | 3.85 | + | - | 4.25 | + | - | 3.45 | + | - | + |

Legend : + (partial) degradation of proteins after 4 hours incubation, large milk clearing zone

- no degradation of proteins after 4 hours incubation, small/no milk clearing zone

Incubation at 30°C: 27 µl medium sample
2.5 µl BSA (25 mg/ml)
0.5 µl Phytase (*A.terreus*, 3-4 g/l)

BSA and phytase were added after the mediumsample was taken from the culture. This sample was incubated at 30°C and after certain timepoints the sample was analysed for the degradation of BSA and phytase.

**Table 5. Protease activity at different pH values**

| STRAINS | Degradation of proteins in Minimal Medium + Trusoy after incubation | | | | | |
|---|---|---|---|---|---|---|
| | pH = 4.5 | | pH = 6 | | pH = 8 | |
| | BSA | Phytase (*A. terreus*) | BSA | Phytase (*A. terreus*) | BSA | Phytase (*A. terreus*) |
| ATCC 9362 | - | - | - | - | - | + |
| ATCC 11906 | - | - | - | - | - | - |
| ATCC 20235 (= *A. oryzae*) | + | - | + | + | + | + |
| ATCC 20387 | - | - | - | - | - | - |

Legend : + (partial) degradation of proteins after 4 hours incubation
- no degradation of proteins after 4 hours incubation

Incubation at 30°C: 25 µl mediumsample → 50mM NaAc pH=4.2
2 µl buffer (50mM) → 50mM NaAc pH=5.8
2.5 µl BSA (25 mg/ml) → 50mM Tris/HCl pH=8.3
0.5 µl Phytase (*A.terreus*, 3-4 g/l)

BSA, phytase and buffer were added after the mediumsample was taken from the culture. This sample was incubated at 30°C and after certain timepoints the sample was analysed for the degradation of BSA and phytase.

**Table 6. PCR results for cloning fungal *pclA* genes**

| Primercombination | | Expected size PCR product |
|---|---|---|
| 1 | pcl1 + pcl2rev | 180 bp |
| 2 | pcl1 + pcl3 | 350 bp |
| 3 | pcl1 + pcl4 | 500 bp |
| 4 | pcl1 + MBL1372 | 300 bp |
| 5 | pcl2 + pcl3 | 200 bp |
| 6 | pcl2 + pcl4 | 350 bp |
| 7 | pcl2 + MBL1372 | 150 bp |
| 8 | MBL1298 + pcl2rev | 180 bp |
| 9 | MBL1298 + pcl3 | 350 bp |
| 10 | MBL1298 + pcl4 | 500 bp |

| Strain | Primercombination | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *Trichoderma reesei* QM9414 | + | ⊞ | - | - | - | - | - | + | - | - |
| *Penicillium chrysogenum* P2 | + | ⊞ | + | - | + | - | - | + | - | - |
| *Fusarium venenatum* ATCC20334 | + | ⊞ | + | - | - | + | - | + | + | - |
| *Trametes versicolor* TV1 | - | - | - | - | - | - | - | - | - | - |
| *Rhizopus oryzae* ATCC200076 | ⊞ | - | - | - | - | - | - | + | - | - |
| *Agaricus bisporus* HORST | - | + | - | - | - | - | - | - | ⊞ | - |
| *Aspergillus sojae* ATCC11906 | + | + | + | - | - | + | - | + | - | - |
| positive control | + | + | + | + | + | + | + | + | + | + |

Legend: + specific PCR product

\- aspecific or no PCR product

⊞ this PCR product was used for sequencing

**Table 7. The viscosity ranges of the various *A. sojae* strains**

| Strains | Viscosity (cP) | | | Biomass (g/l) |
|---|---|---|---|---|
| | Shear rate 6.5 l/s | Shear rate 83.2 l/s | Shear rate 644.4 l/s | |
| *A. sojae* wild type | >>2000 | 1505 | 155 | 8.8 and 16.6 |
| *A. sojae pclA* | 2000 | 751 | 76 | 7.6 and 17.2 |
| *A. sojae lfvA* | 1565 | 94 | 18 | 6.9 and 18.8 |

**Table 8. Promoter strength in *A. sojae* transformants**

| Transformants | Promoter | GUS activity (U/mg) in Minimal Medium | | | | |
|---|---|---|---|---|---|---|
| | | 5% xylose | 5% glucose | 5% maltodextrin | 2% starch | 5% trusoy |
| ATCC11906 wild type | ---- | 0 | 0 | 0 | 0 | 0 |
| ATCC11906[pGUS54] | *gpdA* | 9141 | 6291 | 6667 | 6937 | 3391 |
| ATCC11906[pGUS64] | *glaA* | 33 | 50 | 25 | 51 | 176 |
| ATCC11906[pBIPGUS] | *bipA* | 2914 | 2849 | 1642 | 493 | 2083 |

CLAIMS

1. A recombinant *Aspergillus sojae* comprising an introduced acetamidase S (*amdS*) gene as selectable marker.

2. An *Aspergillus sojae* according to claim 1, said *Aspergillus sojae* being selectable on a medium comprising a substrate for the introduced amdS as sole source of nitrogen, said medium further comprising a carbon substrate and said medium being free of endogenous amdS inducing substrate.

3. An *Aspergillus sojae* according to claim 1 or 2, wherein the source of nitrogen is acrylamide.

4. An *Aspergillus sojae* according to any of the preceding claims wherein the *Aspergillus sojae* has no active endogenous *amdS* gene, for example because the endogenous *amdS* gene comprises an endogenous amdS inactivating mutation, e.g. a deletion or a disruption.

5. A method of introducing a nucleic acid sequence into *Aspergillus sojae,* said method comprising subjecting *Aspergillus sojae* to a method of introduction of a nucleic acid sequence e.g. transformation or transfection of the *Aspergillus sojae* in a manner known per se for introduction of a nucleic acid sequence into fungi, said method comprising the introduction of the *amdS* gene as the nucleic acid sequence (henceforth the introduced *amdS* gene) followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium free of endogenous amdS inducing substrate, said medium further comprising a substrate for the introduced amdS as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired *Aspergillus sojae* comprising the nucleic acid sequence to grow whilst eliminating growth of *Aspergillus sojae* free of the socalled introduced nucleic acid sequence due to inability of such *Aspergillus sojae* to grow without the introduced *amdS* gene on the selection medium, said medium suitably comprising a substrate for amdS other than acetamide, for example acrylamide as substrate for the introduced amdS as sole source of nitrogen.

6. An *Aspergillus sojae* obtained by the method of claim 5.

7. A method of selecting transformed or transfected *Aspergillus sojae* said method

comprising subjecting *Aspergillus sojae* according to any of claims 1-4 and 6 to a method of transformation or transfection of the *Aspergillus sojae* in a manner known per se for transformation or transfection of fungi with a nucleic acid sequence, said method comprising the introduction of the *amdS* gene as the nucleic acid sequence followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium comprising a substrate for the introduced amdS as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired *Aspergillus sojae* to grow whilst eliminating growth of non-transformed or -transfected *Aspergillus sojae* due to inability of such to grow without the introduced *amdS* gene on the selection medium.

8. A method for producing recombinant *Aspergillus sojae*, said method comprising introduction of a nucleic acid sequence into an *Aspergillus sojae* e.g. by transformation or transfection in a manner known per se according to any of claims 1-4 and 6, said nucleic acid sequence comprising a desired sequence to be introduced flanked by sections of an endogenous *amdS* gene or corresponding sequences of a length and homology sufficient to ensure recombination thus simultaneously eliminating the endogenous *amdS* gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for a selectable marker comprised in or transformed in cotransformation with the desired sequence, said selectable marker being absent in the *Aspergillus sojae* prior to introduction of the nucleic acid sequence, suitably the selectable marker being pyrG.

9. An *Aspergillus sojae* exhibiting growth with medium comprising uracil and fluoro-orotic acid, said *Aspergillus* further not exhibiting growth on medium comprising uridine and fluoro-orotic acid, i.e. said *Aspergillus sojae* exhibiting uracil auxotrophy, said *Aspergillus sojae* being unable to utilize uridine, said *Aspergillus sojae* being pyrG negative, said *Aspergillus sojae* exhibiting resistance to fluoro-orotic acid, said uracil auxotrophy and said fluoro-orotic acid resistance being relievable upon complementation with an active introduced *pyrG* gene, suitably said *Aspergillus sojae* being free of active endogenous *pyrG* genes; e.g. the *Aspergillus sojae* endogenous *pyrG* gene comprises a mutation in the form of an insertion, substitution or deletion in the gene or in a gene regulating sequence, e.g. a deletion of the whole coding

sequence of the gene.

10. An *Aspergillus sojae* according to claim 9 in combination with the characteristics of an *Aspergillus sojae* according to any of claims 1-4 and 6.

11. A method of selecting transformed or transfected *Aspergillus sojae*, said method comprising subjecting *Aspergillus sojae* according to claim 9 or 10 to a method of transformation or transfection with a nucleic acid sequence, said method comprising introducing an active *pyrG* gene into the *Aspergillus sojae* in a manner known per se for transformation or transfection of fungi followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium free of uracil and fluoro-orotic acid, said medium at least further comprising minimum substrates required for growth of *Aspergillus sojae*, said medium enabling the desired *Aspergillus sojae* to grow whilst eliminating growth of non-transformed or -transfected *Aspergillus sojae* due to inability of such to grow without uracil due to the inactivated *pyrG* gene.

12. A method according to claim 11, wherein the active *pyrG* gene that is introduced is flanked by identical nucleic acid sequence fragments, and the pyrG positive *Aspergillus sojae* resulting from the introduction of the *pyrG* gene and the flanking sequences is selected on a medium free of uracil and fluoro-orotic acid and subsequently the pyrG positive *Aspergillus sojae* is cultivated on medium comprising uracil and fluoro-orotic acid thereby eliminating the *pyrG* gene that had been introduced thus resulting in a pyrG negative *Aspergillus sojae* that is selectable by growth on uracil comprising medium and by fluoro-orotic acid resistance, suitably the flanking sequences and the *pyrG* gene being further flanked by sequences that direct integration of the *pyrG* gene and the flanking sequences into a specific location due to the fact that the integration directing sequences are homologous to a specific sequence of the *Aspergillus sojae* to be transformed, thereby enabling knock out, if desired, of the gene associated with the specific sequence.

13. A method according to claim 11 or 12, wherein the *Aspergillus sojae* according to claim 9 or 10 has a further nucleic acid sequence introduced therein, preferably said further nucleic acid sequence encoding a protein or polypeptide, said further nucleic acid sequence being introduced with the active *pyrG* gene either on the same vector or by cotransformation with the active *pyrG* gene that is introduced.

14. A method of selecting transformed or transfected *Aspergillus sojae* by carrying out the method according to any of claims 11-13 in combination with the method of claim 5.

15. A method for producing recombinant *Aspergillus sojae*, said method comprising introducing a nucleic acid sequence into a pyrG positive *Aspergillus sojae*, e.g. by transformation or transfection in a manner known per se, said nucleic acid sequence comprising the desired sequence flanked by sections of the *pyrG* gene or corresponding sequences of a length and homology sufficient to ensure recombination eliminating the *pyrG* gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for *Aspergillus sojae* with a pyrG negative phenotype.

16. A recombinant *Aspergillus sojae* obtained by a method according to any of claims 11-15, optionally further comprising the characteristics of an *Aspergillus sojae* according to any of claims 1-4, 6, 9 and 10.

17. A recombinant *Aspergillus sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being susceptible to degradation upon expression by *Aspergillus niger* or *Aspergillus awamori*.

18. A recombinant *Aspergillus sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being other than *Aspergillus sojae* protease and amylase, said protein or polypeptide preferably being a non-*Aspergillus sojae* protein or polypeptide.

19. A mutant or recombinant *Aspergillus sojae* comprising a mutation inactivating a protease gene, suitably an alkaline protease gene.

20. A mutant or recombinant *Aspergillus sojae* comprising a mutation inactivating the major protease gene, suitably a mutation inactivating the major alkaline protease gene, e.g. the gene encoding major alkaline protease gene of 35 kDa.

21. A method for producing recombinant *Aspergillus sojae*, said recombinant *A. sojae* exhibiting reduced proteolytic activity, said method comprising introduction into an *A. sojae*, e.g. by transformation or transfection in a manner known per se, of a nucleic acid sequence comprising a selectable marker encoding sequence to be introduced flanked by sections of the protease gene to be eliminated and further said flanking

sequence and the selectable marker encoding sequence being comprised within sequences of a length and homology sufficient to ensure recombination at the protease gene thus simultaneously eliminating the protease gene and introducing the desired selectable marker encoding sequence, the introduction being followed by selection of the recombinant *A. sojae* by selecting for the selectable marker, whereby the *A. sojae* prior to the introduction of the nucleic acid sequence, e.g. by transformation or transfection, is free of the selectable marker to be introduced, e.g. the *A. sojae* prior to the introduction of the nucleic acid sequence being mutated such that the *A. sojae* cannot produce active selectable marker, suitably the selectable marker being the *pyrG* gene, suitably the method being carried out together with the method according to any of claims 11-15.

22. A recombinant *Aspergillus sojae* obtained according to the method of claim 21.

23. A mutant or recombinant *Aspergillus sojae* according to any of claims 17-20 or 22 comprising a selectable marker, preferably amdS as defined in any of claims 1-4 or 6 and/or pyrG as defined in claims 9, 10 or 16.

24. A recombinant *Aspergillus sojae* according to any one of claims 1-4, 6, 9, 10, 16-20, 22 and 23, comprising an introduced nucleic acid sequence encoding phytase or a protein having phytase activity.

25. A process of expression of an introduced nucleic acid sequence encoding a protein or polypeptide comprised in a recombinant or mutant *Aspergillus sojae* as defined in any of the claims 1-4, 6, 9, 10, 16-20, 22-24 or obtained via a method according to any of claims 5, 7, 8, 11-15 and 21, said process comprising cultivating the recombinant or mutant *A. sojae,* suitably the introduced nucleic acid sequence encoding a protein or polypeptide being absent in the corresponding non-transformed or wild-type *A. sojae* and/or being present in a lower copy number.

26. A recombinant fungus comprising a mutation in a gene encoding a proprotein convertase or a functionally equivalent protein.

27. A fungus according to claim 26 exhibiting increased production of a protein, polypeptide or metabolite under equivalent conditions when compared to the corresponding wild-type fungus.

28. A fungus according to claims 26 or 27 said mutation being obtained by specific gene

modification using transformation or transfection in a manner known per se.

29. A fungus according to claims 26-28, said proprotein convertase or functionally equivalent protein being encoded by a nucleotide sequence of which a fragment can be amplified by in vitro DNA amplification using any of two mixtures of nucleotides given in SEQ ID Nos. 10 to 16.

30. A fungus as described in claim 27, said proprotein convertase or functionally equivalent protein being encoded by a nucleotide sequence allowing functional complementation of the growth phenotype of an *Aspergillus niger* mutant comprising a mutation which inhibits the activity of a proprotein convertase or a functionally equivalent protein.

31. A fungus according to any of claims 26-30, said fungus being an *Aspergillus sojae.*

32. A fungus according to any of claims 26-30, said fungus further containing an introduced *amdS* gene or *pyrG* gene.

33. A process for expressing a protein or polypeptide, preferably a recombinant protein or polypeptide, encoded by a nucleotide sequence, said process comprising cultivating a fungus according to any of the claims 26-32.

34. A process for producing a protein or polypeptide, preferably a recombinant protein or polypeptide, said process comprising a process of expression according to claim 33, optionally including processing and/or secretion and/or isolation of the expressed protein or polypeptide.

35. A process for producing a phytase or a protein having phytase activity, preferably a recombinant phytase or recombinant protein having phytase activity, said process comprising a process of expression according to claim 33, optionally including processing and/or secretion and/or isolation of the expressed phytase or protein having phytase activity.

Fig 1 (1)

1 50

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 ---------- ---------- ---------- ---------- ----------

KLKEX1 ---------- ---------- ---------- ---------- ----------

CAKEX2 ---------- ---------- ---------- ---------- ----------

SPKRP ---------- ---------- ---------- ---------- ----------

YLKEX2 MLRKFILGLL LASQAVAQLP HKERDYDSRV YVALSLRDGL DPREFEASVS

51 100

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 ---------- ---------- ---------- ---------- ----------

KLKEX1 ---------- ---------- ---------- ---------- ----------

CAKEX2 ---------- ---------- ---------- ---------- -MLPIKLLIF

SPKRP ---------- ---------- ---------- ---------- ----------

YLKEX2 GLDHGQWTFE HPVGTIPNTY VFSAPKEYAP IENIRDQDRL EVAGGVLAKR

101 150

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

Fig 1 (2)

SCKEX2 ---------- ---------- ---------M KVRKYITLCF WWAFSTSALV

KLKEX1 ---------- ---------- ---------- --MILSSQLM LALIAVSGYG

CAKEX2 ILGYLLSPTL QQYQQIPPRD YENKNYFLVE LNTTNSQKPL IDFISHYRGH

SPKRP ---------- ---------- ---------- ---------- -------MHP

YLKEX2 ELRKREKLQK KYGMSEEDVE KRLVALERLD YDWSERGLGS LEVLSERRIH

151 200

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 SSQQIPLKDH TSRQYFAVES NETLSRLEEM HPNWKYEHDV RGLPNHYVFS

KLKEX1 KAMQVPKKDH ENRQYFAIES YDDVGNLLAE HSDWSFEHDV RGLANHYVFS

CAKEX2 YNFEHQLSSL DNHYVFSIDK SHPHNSFLGN HNSNEYNLMK RQLGHEQDYD

SPKRP ALLCGPILAI FLQFLVSSCS PLENDDLFLV QVEPEVDPVV AAEAIGAKYV

YLKEX2 KRAPVNWTEE EMEYLKEIKR RAEEAQKAQD DKGDKKEDQK DDKKEGQEAQ

Fig 1(3)

201 250

XENPC2 ---------- ---------- -KVNKVEQQE GFHRKKKR.. ..........

XNFURIN ---------- ---------- -QVHWLEQQV AKKRKKKR.. ..........

SCKEX2 KELLKLGKRS SLEELQGDNN DHILSVHDLF PRNDLFKRLP VP.....APP

KLKEX1 KPLQSLGKRD AIDTGYSEN. ..IIDFHDL. PPVQLHKRLP IG.......

CAKEX2 ELISHVESIH LLPMKKLSKR IPVPIEMEDV VFDNRDDTGS DN.....HEA

SPKRP RPLLNLKYHH LIKLHKGSDD SVQSSIRKRG IDAGILELER QTPRWRYKRD

YLKEX2 KEGDKEDNKG DDKEDGEEDD DDDEDEDDDD ASPAMPVQWK PVDESMYGGM

251 300

MBL793

XENPC2 .....GYR.. DINDIEINMN DPLFTKQWYL INTGQDADGT PGLDLNVAEA

XNFURIN .......... ...DIYTDPT DPKFMQQWYL LDTNRH.... ...DLHVKEA

SCKEX2 MDSSLLPV.. KEAEDKLSIN DPLFERQWHL VNPSFPGS.. ...DINVLDL

KLKEX1 .DSSMEQI.. QNARILFNIS DPLFDQQWHL INPNYPGN.. ...DVNVTGL

CAKEX2 TDEAHQKL.. IEIAKKLDIH DPEFTTQWHL INLKYPGH.. ...DVNVTGL

SPKRP ASESDELL.. NEFSNHFGIS DPLFYGQWHI FNSNNPGH.. ...DLNLREV

YLKEX2 PDDSLYDVYR KYYPDEVGIK DPSLWKQWYL HNVHKAGH.. ...DLNVTGL

Fig 1(4)

301 350

MBL1208\ /MBL794

XENPC2 WE1GYTGRGV TIAIM.DDGI DYLHPDLASN YNAEASYDFS SNDPYPYPRY

XNFURIN WEQGFTGKGI VVSILSDDGI EKNHPDLQAN YDPAASYDVN DQDPPPQPKY

SCKEX2 WYNNITGAGV VAAIV.DDGL DYENEDLKDN FCAEGSWDFN DNTNLPKPRL

KLKEX1 WKENITGYGV VAALV.DDGL DYENEDLKDN FCVEGSWDFN DNNPLPKPRL

CAKEX2 WLEDILGQGI VTALV.DDGV DAESDDIKQN FNSEGSWDFN NKGKSPLPRL

SPKRP WDAGYFGENV TVAFV.DDGI DFKHPDLQAA YTSLGSWDFN DNIADPLPKL

YLKEX2 WLRNVTGWGV VTAVV.DDGL DMNAEDIKAN YFAEGSWDFN FNKSDPKPSS

351 400

/MBL1158

PE4,PCL1\

XENPC2 TDDWFNSHGT RCAGEVSASA NNNICGVGVA YNSKVAGIRM LDQPFMTDII

XNFURIN TQLNDNRHGT RCAGEVAAVA NNGICGVGIA YNANIGGVRM LDGE.VTDAV

SCKEX2 SDDY...HGT RCAGEIAAKK GNNFCGVGVG YNAKISGIRI LSGD.ITTED

KLKEX1 KDDY...HGT RCAGEIAAFR .NDICGVGVA YNSKVSGIRI LSGQ.ITAED

CAKEX2 FDDY...HGT RCAGEIAAVK .NDVCGIGVA WKSQVSGIRI LSGP.ITSSD

SPKRP SDDQ...HGT RCAGEVAAA. WNDVCGVGIA PRAKVAGLRI LSAP.ITDAV

YLKEX2 HDDY...HGT RCAGEIAAVR .NNVCGVGVA YDSKVAGIRI LSKE.IAEDI

5/45

Fig 1 (5)

401 450

/PCL2(rev)\

XENPC2 EASSISHMPQ VIDIYSASWG PTDDGKTVDG PRELTLQAMA DGVNKGRGGK

XNFURIN EARSLGLNPN HIHIYSASWG PEDDGKTVDG PAKLAEEAFY RGVTQGRGGL

SCKEX2 EAASLIYGLD VNDIYSCSWG PADDGRHLQG PSDLVKKALV KGVTEGRDSK

KLKEX1 EAASLIYGLD VNDIYSCSWG PSDDGKTMQA PDTLVKKAII KGVTEGRDAK

CAKEX2 EAEAMVYGLD TNDIYSCSWG PTDNGKVLSE PDVIVKKAMI KGIQEGRDKK

SPKRP ESEALNYGFQ TNHIYSCSWG PADDGRAMDA PNTATRRALM NGVLNGRNGL

YLKEX2 EALAINYEMD KNDIYSCSWG PPDNGQTMAR PGKVVKDAMV NAITNGRQGK

451 500

/ PE6 / PCL3 /MBL789

XENPC2 GSIYVWASGD GG.SYDDCNC DGYASSMWTI SINDAINDGR TALYDESCSS

XNFURIN GSIYVWASGN GGREHDSCNC DGYTNSIYTL SISSTTQMGN VPWYSEACSS

SCKEX2 GAIYVFASGN GGTRGDNCNY DGYTNSIYSI TIGAIDHKDL HPPYSEGCSA

KLKEX1 GALYVFASGN GGMFGDSCNF DGYTNSIFSI TVGAIDWKGL HPPYSESCSA

CAKEX2 GAIYVFASGN GGRFGDSCNF DGYTNSIYSI TVGAIDYKGL HPQYSEACSA

SPKRP GSIFVFASGN GGHYHDNCNF DGYTNSIFSA TIGAVDAEHK IPFYSEVCAA

YLKEX2 GNVFVFASGN GGSRGDNCNF DGYTNSIYSI TVGALDFNDG HPYYSEACSA

6/45

Fig 1(6)

501 550

/ PCL4

XENPC2 TLASTFSNGR KRNPEAGVAT TDLY..... ..GNCTLRHS GTSAAAPEAAG

XNFURIN TLATTYSSGN QN..EKQIVT TDLR..... ..QKCTDSHT GT---------

SCKEX2 VMAVTYSSGS GEY....IHS SDIN..... ..GRCSNSHG GTSAAAPLAAG

KLKEX1 VMVVTYSSGS GNY....IKT TDLD..... ..EKCSNTHG GTSAAAPLAAG

CAKEX2 VMVVTYSSGS GEH....IHT TDI..... .KKKCSATHG GTSAAAPLASG

SPKRP QLVSAYSSGS HLS....ILT TN.P..... .EGTCTRSHG GTSAAAPLASA

YLKEX2 NMVVTYSSGS EHY....IVG TDINAIDDKS AAPRCQNQHG GTSAAAPLAAG

551 600

/MBL1219

XENPC2 VFALALEANP GLTWRDLQHL SVLTSK---- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 VYTLLLEANP NLTWRDVQYL SILSAVGLE. KNADGDWRDS AMGKKYSHRY

KLKEX1 IYTLVLEANP NLTWRDVQYL SILSSEEIN. PH.DGKWQDT AMGKRYSHTY

CAKEX2 IYSLILSANP NLTWRDVQYI SVLSATPIN. EE.DGNYQTT ALNRKYSHKY

SPKRP VYALALSIRP DLSWRDIQHI TVYSASPFDS PSQNAEWQKT PAGFQFSHHF

YLKEX2 VFALALSVRP DLTWRDMQYL ALYSAVEIN. .SNDDGWQDT ASGQRFHHQF

Fig 1(7)

```
601 650

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 GFGKIDAHKL IEMSKTWENV NAQTWFYLPT LYVSQSTNST ..........

KLKEX1 GFGKLDAYNI VHMAKSWINV NPQGWLYLPT IVEKQSISNS ..........

CAKEX2 GYGKTDAYKM VHFAKTWVNV KPQAWYYSDI IEVNQTITTT PEQKAPSKRD

SPKRP GFGKLDASKF VEVAKDWQVV NPQTWLIAPE INVNKSFGSV NNETITE...

YLKEX2 GYGKLDASKI VELAEGWNLV NNQTSFHSEV KTVSQKV... ..........
```

```
651 700

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 ..EETLESVI TISEKSLQDA NFKRIEHVTV TVDIDTEIRG TTTVDLISPA

KLKEX1 ..DEVIESTV SVSAEEFKQN NLKRLEHVTV TVDIDAPYRG HVLVDLISPD

CAKEX2 SPQKIIHSSV NVSEKDLKIM NVERVEHITV KVNIDSTYRG RVGMRIISPT

SPKRP .....MVSEF TVTKDMIEKS NFKRLEHVTV RVCIPFNRRG ALEILLESPS

YLKEX2 KYNEPLKSVI TVTRDDLDKV NFKRAEHITA VLNLEASYRG HVRVLLKGPR
```

Fig 1(8)

```
701 750

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 GIISNLGVVR PRDVSSEGFK DWTFMSVAHW GENGVGDWKI KV...KTTEN

KLKEX1 GVTSTLATAR RLDKNRYGFQ NWTFMSVAHW GSSGVGSWKL KV...KSTHD

CAKEX2 GVISDLATFR VNDASTRGFQ NWTFMSVAHW GETGIGEWKV EVFVDDSKGD

SPKRP GIRSILASER PYDENSKGFL DWTFMTVQHW AEPPEGVWKL ..LVNDRSGG

YLKEX2 GVVSELAALR RDDRSKDGYD NWAFMSVAHW ADEGEGDWEL TV...ENTGE

751 800

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 GHRIDFHSWR LKLFGESIDS SKTETFVFGN DKEEVEPAAT ESTVSQYSAS

KLKEX1 NEIVTLKSWR LKMFGETIDA KKAKVISYGN DKEDAEVKST E.......S

CAKEX2 QVEINFKDWQ FRIFGESIDG DKAEVYDITK DYAAIR...R ELLEKEKQNS

SPKRP KHEGTFENWQ LALWGESENP SNTAPLPYDT LELPKEMVLG IYSEPNSDLT

YLKEX2 QDQVELVNWQ LNVFGEQKDK REENKEGESK PEDENKEGEK EGEKKPEDEN
```

Fig 1(9)

```
801 850

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 STSISISATS TSSISIGVET SA..IPQTTT ASTDPDSDPN TPKKLSSPRQ

KLKEX1 KTTTPTAQTS SFTTTSGEET SG..... ..... .ANKLPRPEQ

CAKEX2 KSTTTTSSTT TATTTSGGEG DQ..KTTTSA ENKESTTKVD N.SASITTSQ

SPKRP NSSTLLSPTS TSFTSYTVSA TA..TPTSTS HIPIPTVLPP TQPVLEPSYR

YLKEX2 KEEGNKEDDK GDQKEDKPED KPEDKPEDTP EDKPEDKPED APEDKPSDEK

851 900

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 AMHYFLTIFL IGATFLVLYF MFFMKSRRRI RRSRAETYEF DIIDTDSEYD

KLKEX1 AAQLYLAIFV IGAIVIIIYY LFFLKSRRII RRSRAEAYEF DIIDTDSEYD

CAKEX2 TASLTSSNEQ HQPTESNSDS DSDTDDENKQ EGEEDNDNDN DNGNKKANSD

SPKRP EIVAFITFFL LFAFIFVAVI WTWISAFWKA KAPPPLSQQE IA* ------

YLKEX2 KPEEKPEEKP VDNSDSSSDS SDSHTSWWPD LSSKKSAWLY GAVLLVGGFI
```

Fig 1(10)

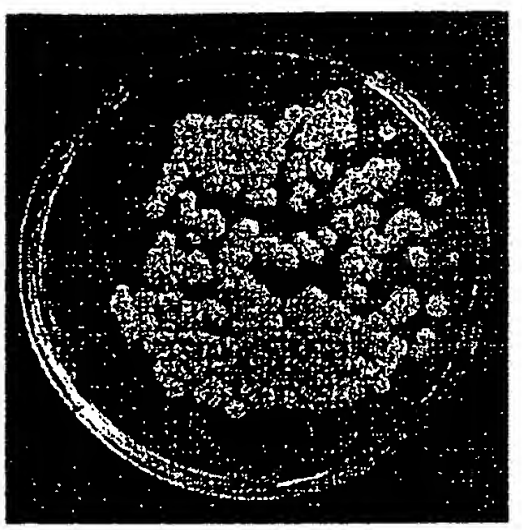
```
901 950

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 STLDNGTSGI TEPEEVEDFD FDLSDEDHLA SLSSSENGDA EHTIDSVLTN

KLKEX1 ASINKLQS-- ---------- ---------- ---------- ----------

CAKEX2 NTGFYLMSIA VVGFIAVLLV MKFHKTPGSG RRRRRRDGYE FDIIPGEDYS

SPKRP ---------- ---------- ---------- ---------- ----------

YLKEX2 AVIGIYACVT RRNRVRRNRS KDAPSASSFE FDLIPHDDSD DDFVYPEDTH

951 1000

XENPC2 ---------- ---------- ---------- ---------- ----------

XNFURIN ---------- ---------- ---------- ---------- ----------

SCKEX2 ENPFSDPIKQ KFPNDANAES ASNKLQELQP DVPPSSGRS* ----------

KLKEX1 ---------- ---------- ---------- ---------- ----------

CAKEX2 DSDDDEDDSD TRRADDDSFD LGHRNDQRVV SASQQQRQYD RQQDEARDRL

SPKRP ---------- ---------- ---------- ---------- ----------

YLKEX2 RRSGDNDRLY DPFAEVEDDD DMFRISDEGE DAHDVEPELN RVSMEADKRD
```

Fig 1(11)

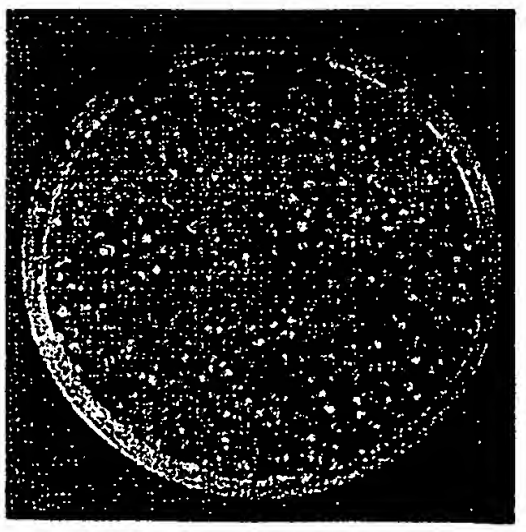
```
1001 1053

XENPC2 ---------- ---------- ---------- --------- ---------- ---

XNFURIN ---------- ---------- ---------- --------- ---------- ---

SCKEX2 ---------- ---------- ---------- --------- ---------- ---

KLKEX1 ---------- ---------- ---------- --------- ---------- ---

CAKEX2 FDDFNAESLP DYENDMFKIG DEEEEEEEEE EGQQSAKAP SNSEGNSGTS TKK

SPKRP ---------- ---------- ---------- --------- ---------- ---

YLKEX2 NDRQNLLG*- ---------- ---------- --------- ---------- ---
```

Fig 2a

Non-selection medium:



Selection medium:
(WO97/04108)



Improved acrylamide
selection medium:

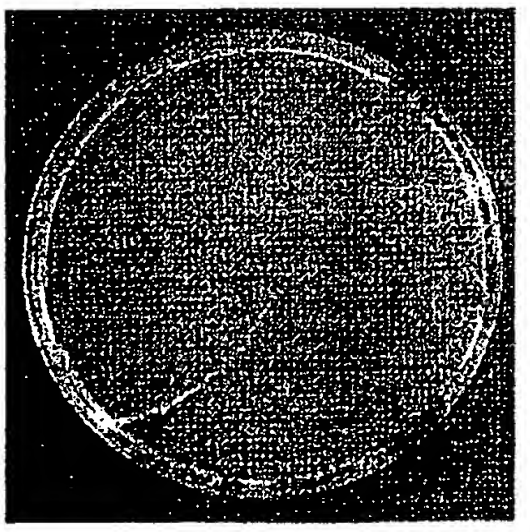
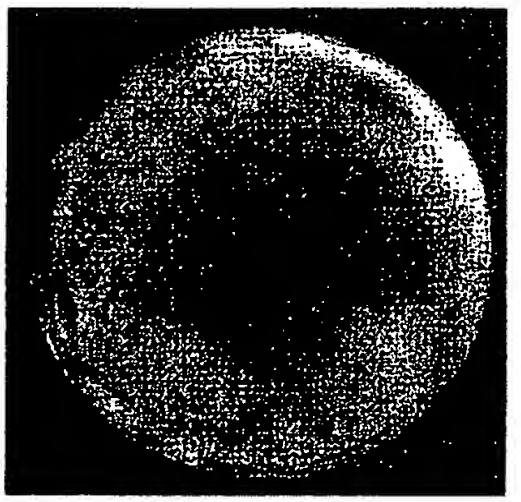
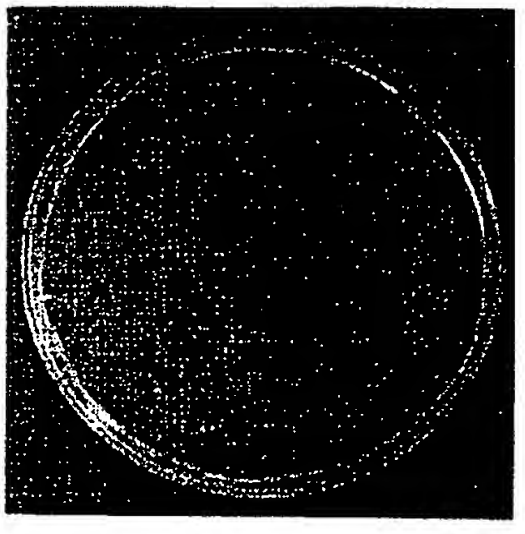


Fig 2b

Non-selection medium:



Selection medium:
(WO97/04108)



Improved acrylamide
selection medium:

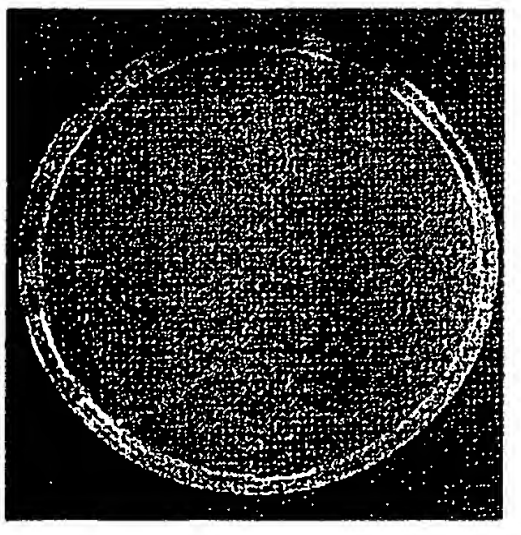
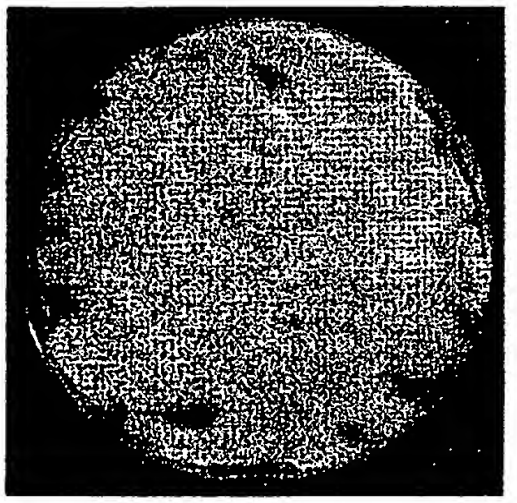
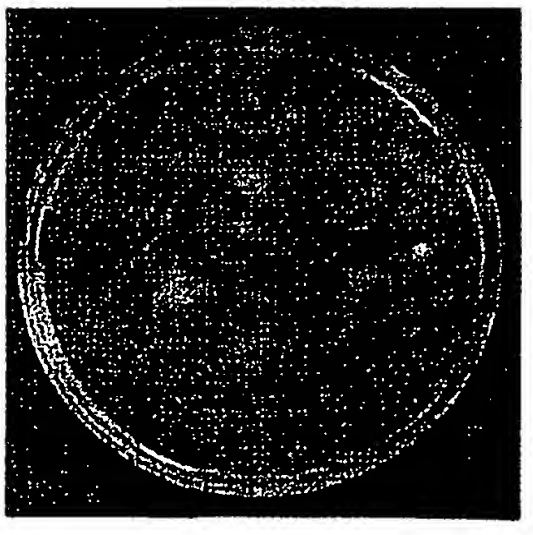


Fig 2c

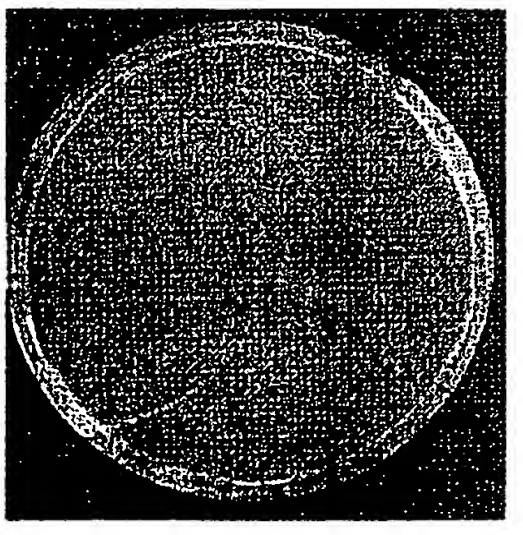
Non-selection medium:



Selection medium:
(WO97/04108)



Improved acrylamide
selection medium:



# Fig 3a

A.oryzae 1 GTCAGTCCCA ATAGAAGGCT CGGTCATCGA TCTACCTGAG AAGTCTGGGA

ATCC11906 GTCAGTCCCA ACAGAAGGCT CGGTCATCGA TCTACCTGAG AAGTCTGGGA

A.oryzae 51 TTCTGTCGCC TTCTGAAATA AAGATTACAA ACTCGTCTGC CACAGAACTT

ATCC11906 TTCTGTCATC TTCTGAAATA AAGATTACAA ATTCGTCTGC CACAGAACTT

A.oryzae 101 GTCGCTCAAT TAGCCAATGG CACGTTGAAG TCCGTGGATG TGACACTCGC

ATCC11906 GTCGCTCAAT TAGCCAATGG CACGTTGAAG TCCGTAGACG TGACACTGGC

INTRON I

A.oryzae 151 ATTCTGTAAA AGAGCTGCAC TGGCTCATCA ACTTgtgggt ataacct---

ATCC11906 ATTCTGTAAA AGAGCTGCAC TGGCTCATCA ACTTgtgagt ataacttgc-

# Fig 3b

A.oryzae 1 -----CCCAA GGCAGATTTG AATGAGGTAT GGGACGCGCA GCTGCAAAAA

ATCC11906 ---------- ---------- ----------T GGGACGCGCA GCTGCAAAAA

A.oryzae 51 TGGCGTTATC AGTGTGAATA CCTTGACAAG TGGCGCGAAT GGGAGGAACG

ATcc11906 TGGCGTTATC AGTGTGAATA CCTTGACAAG TGGCGCGAAT GGGAGGAACG

A.oryzae 101 GACGGGCAAG GAGCTTGACG CTATCATCGC CCCGGTGGCG GCGACAGCTG

ATCC11906 AACGGGCAAG GAGCTTGACG CTATCATGCC CCCGGTGGCG GCGACAGCTG

A.oryzae 151 CAGTCCGC

ATCC11906 CAGTCCGC

Fig 4a

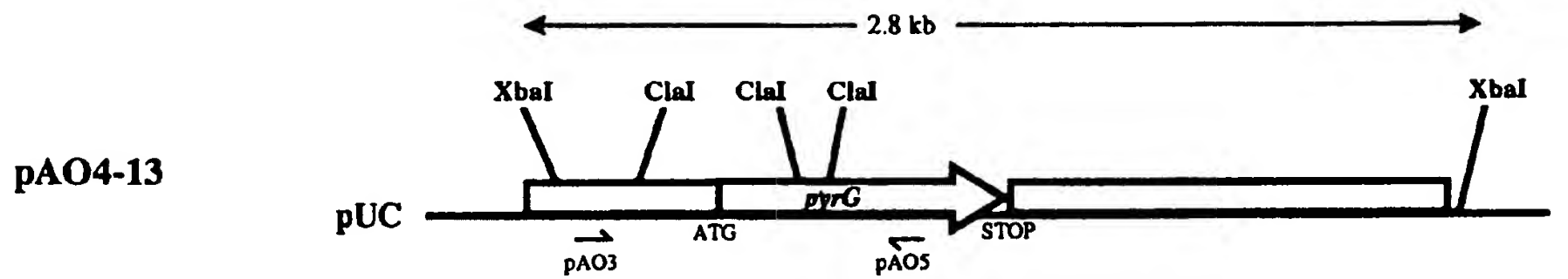

pAO4-13
2.8 kb
XbaI
ClaI
ClaI
ClaI
XbaI
pUC
pyrG
ATG
STOP
pUC
pAO3
pAO5

Fig 4b

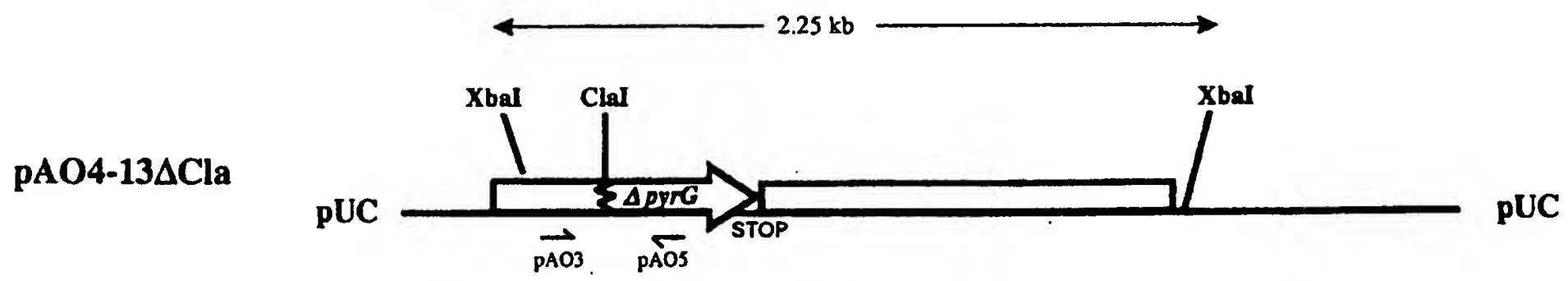
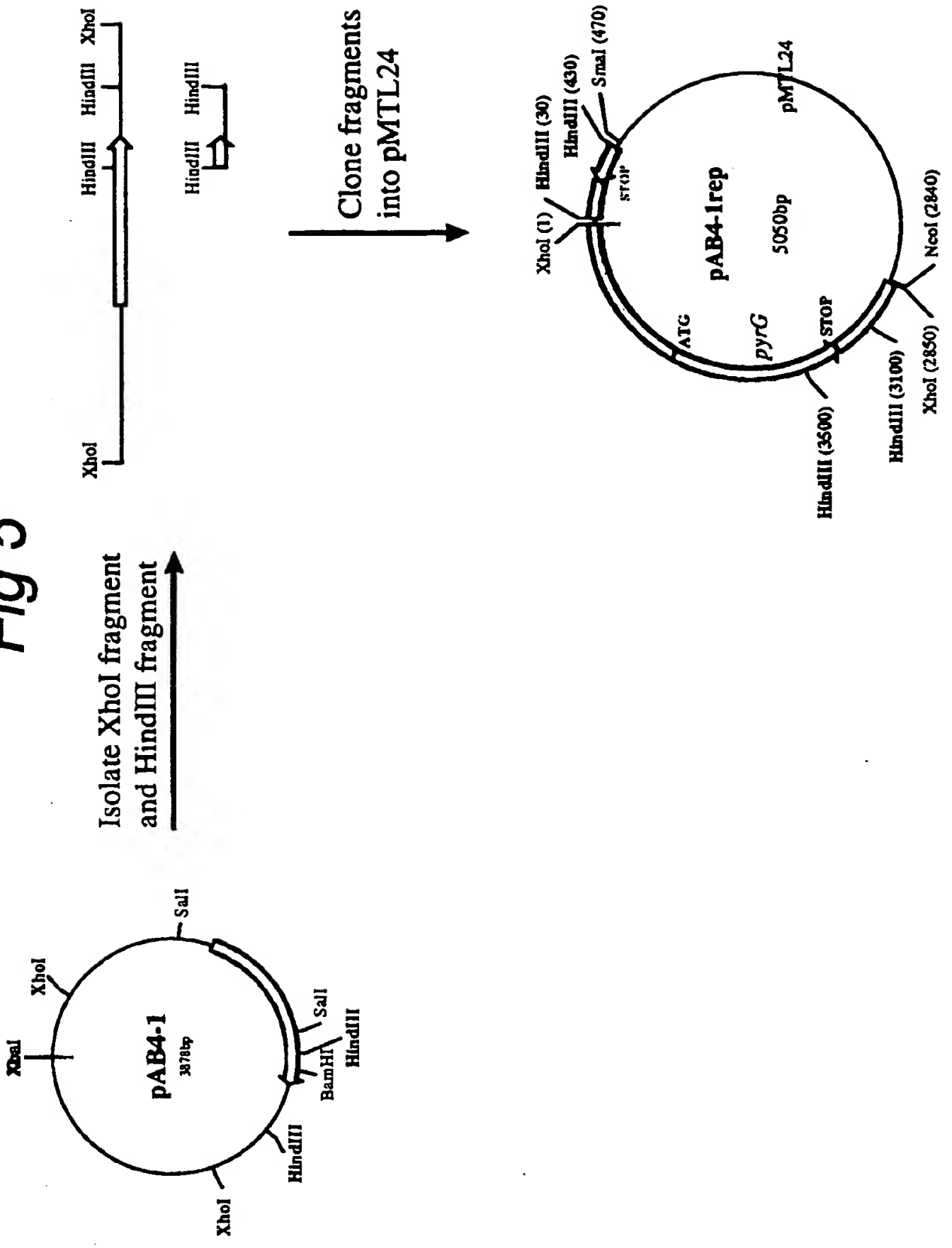
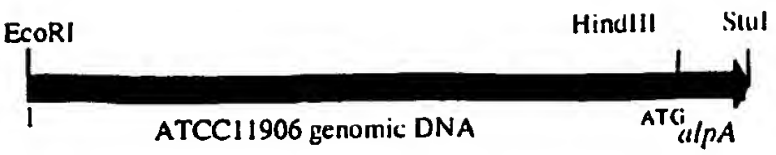

pAO4-13ΔCla
2.25 kb
XbaI
ClaI
XbaI
pUC
Δ pyrG
STOP
pUC
pAO3
pAO5

Fig 5

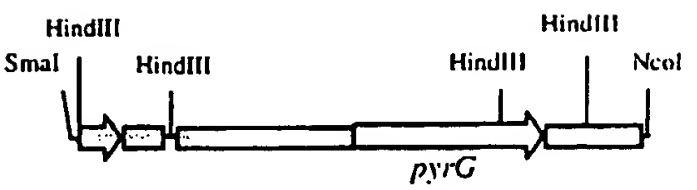

Isolate XhoI fragment and HindIII fragment
XhoI
HindIII
HindIII
HindIII
HindIII
Clone fragments into pMTL24
pAB4-1rep
5050bp
pMTL24
XhoI (1)
HindIII (30)
HindIII (430)
SmaI (470)
NcoI (2840)
XhoI (2850)
HindIII (3100)
HindIII (3500)
ATG
pyrG
STOP
STOP
pAB4-1
3878bp
XhoI
SalI
XhoI
HindIII
SalI
BamHI
HindIII
XhoI

# Fig 6

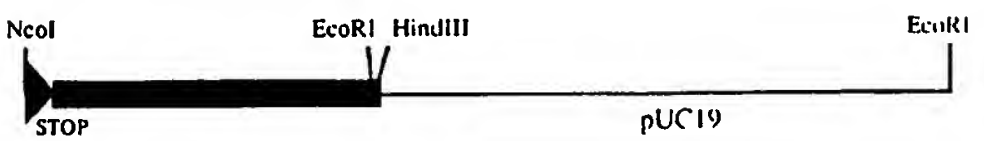
4.4 kb EcoRI-StuI fragment from pAS1-1




2.6 kb SmaI-NcoI fragment from pAB4-1rep




4.4 kb NcoI-EcoRI fragment from pAS1-2A




3-way ligation

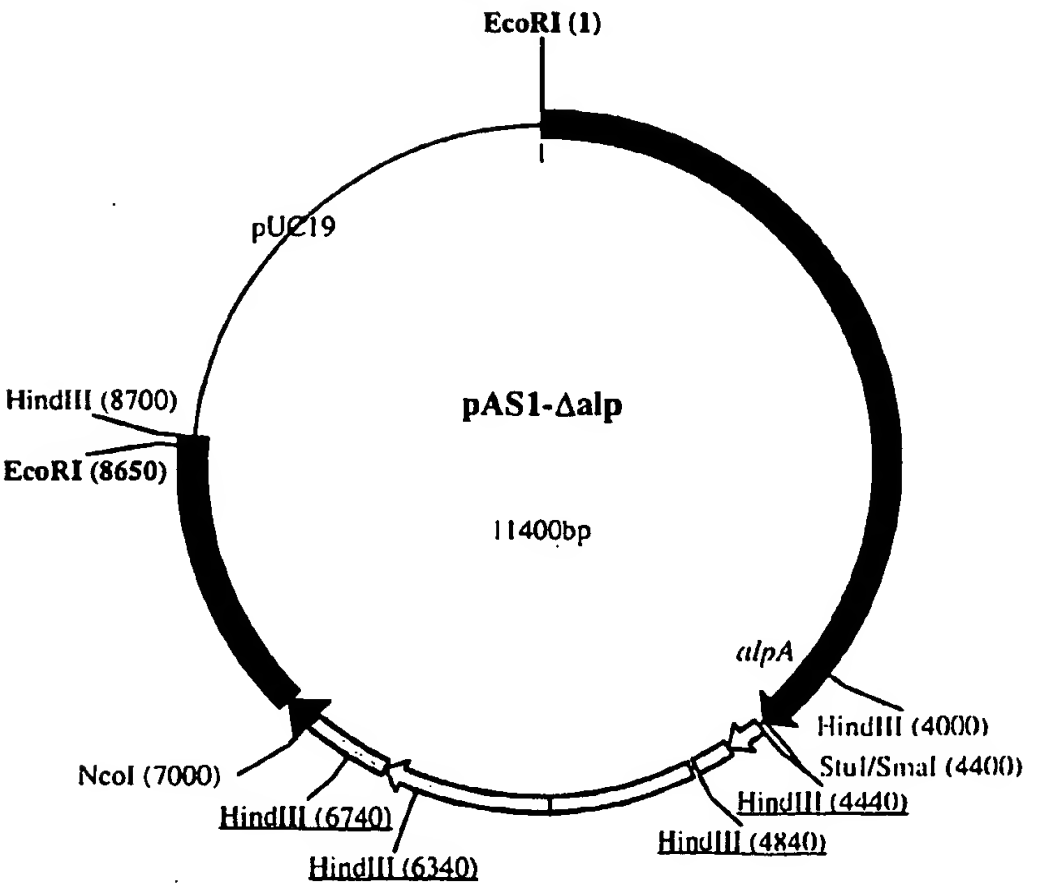


Fig 7

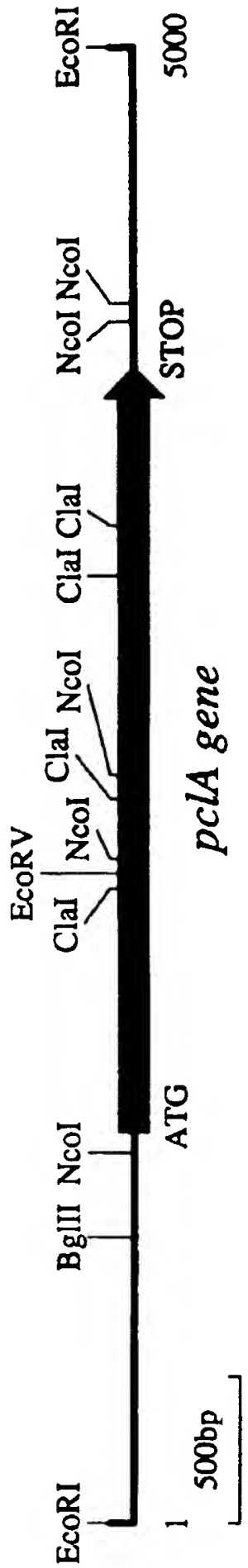
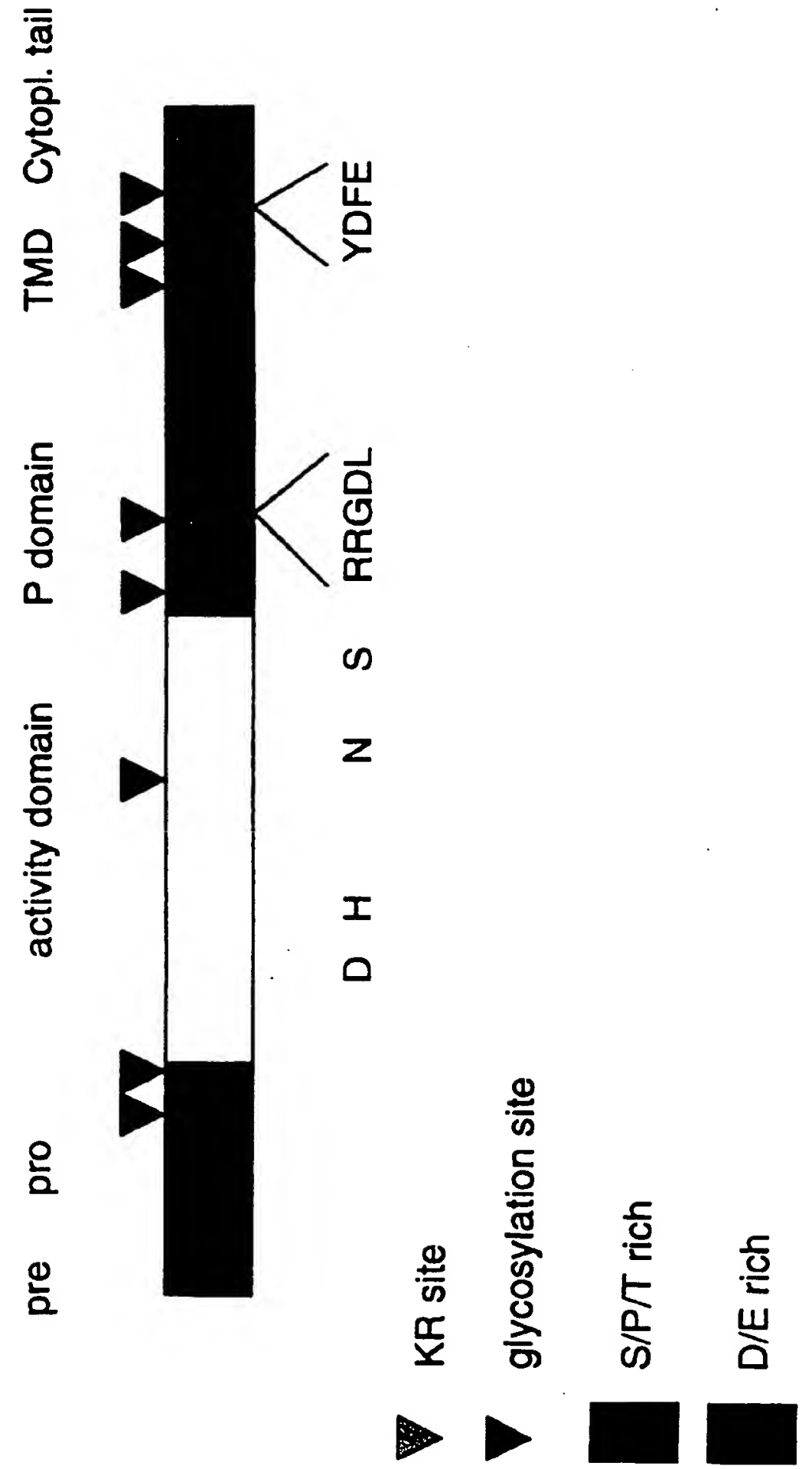

EcoRI
5000
NcoI NcoI
STOP
ClaI ClaI
NcoI
ClaI
EcoRV
NcoI
ClaI
pclA gene
BglII NcoI
ATG
EcoRI
1
500bp

Fig 8


pre
pro
activity domain
P domain
TMD
Cytopl. tail
N
H
D
S
RRGDL
YDFE
KR site
glycosylation site
S/P/T rich
D/E rich

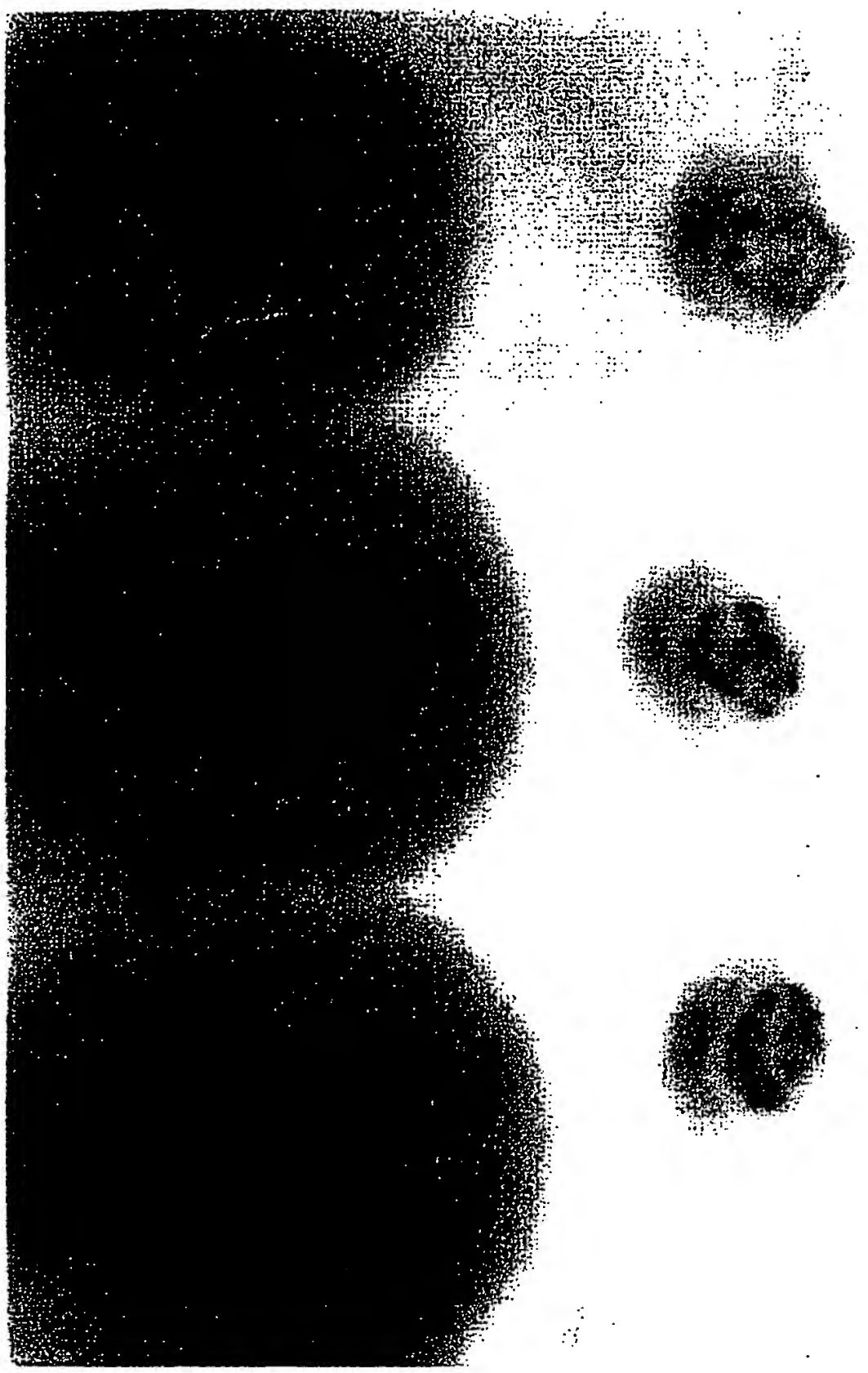



Fig 9

Fig 10 (1)

```
. . . . .

797 CCACGGTACTAGATGCGCGGGTGAAATCGGTGCGGCGAAGAACGACGTGT 846

||| || || |||| |||||||||| || || || | ||| || || |

8 CCATGGAACACGATGTGCGGGTGAAATTGGAGCAGCTAGGAATGATGTCT 57

. . . . .

847 GCGGGGTTGGTGTTGCGTATGATAGTCGCATCGCTGGTATTCGGATTCTC 896

| || || |||||||| || || || | | || || ||||||||| |

58 GTGGAGTAGGTGTTGCATACGACAGCCAAGTTGCCGGAATTCGGATTTTG 107

. . . . .

897 TCCGCACCCATCGATGACACTGATGAGGCTGCGGCTATTAACTACGCCTA 946

|||||||||| || ||| | |||||||||| || || ||||| | ||

108 TCCGCACCCATTGACGACGCAGATGAGGCTGCTGCCATCAACTATGGCTT 157

. . . . .
```

Fig 10 (2) 24/45

```
947 TCAGGAGAACGATATCTACTCGTGTTCCTGGGGTCCCTATGATGATGGCG 996

||| || ||||| || || || ||||||| || ||||||||||

158 CCAGCGCAATGATATATATTCATGCTCCTGGGGCCCTCCGGATGATGGCG 207

. . . . .

997 CCACAATGGAAGCCCCGGGCACTCTGATCAAGCGGGCCATGGTCAATGGT 1046

|||| ||||| || || || | ||| ||||| || || |||||||| |||

208 CCACGATGGAGGCGCCAGGGATTCTTATCAAACGAGCTATGGTCAACGGT 257

. . . . .

1047 ATCCAAAATGGTCGAGGTGGAAAAGGCTCGGTTTTTGTATTTGCGGCTGG 1096

|||||||||| ||||| || ||||| || | || || ||||| |||||

258 ATCCAAAATGGCCGAGGAGGTAAAGGTTCTATCTTCGTCTTTGCAGCTGG 307

. . . . .

1097 TAACGGTGCCATTCATGACGATAACTGTAACTTTGACGGTTACACCAACA 1146

|| ||||| | || || ||||| || || ||||||| || ||||

308 AAATGGTGCAGGGTACGATGACAACTGCAATTTCGACGGTTATACAAACA 357

. . . . .
```

Fig 10 (3) 25/45

```
1147 GTATCTACAGCATCACGGTGGGTGCCATTGATCGGGAGGGTAACCATCCT 1196

|  ||  ||||||||||  ||  ||  ||  |||||||  |||||  ||  |||||

358 GCATTTACAGCATCACCGTCGGCGCTATTGATCGAGAGGGCAAACATCCC 407

.  .  .  .  .

1197 CCGTATTCGGAATCCTGCTCGGCGCAACTGGTGGTTGCCTACAGCAGCGG 1246

||  ||||||||  |||||  ||  ||  ||||  ||  ||  ||  |||||  ||

408 AGCTACTCGGAATCATGCTCTGCCCAGTTGGTTGTCGCTTATAGCAGTGG 457

.  .  .  .  .

1247 CGCCAGTGATGCAATTCATACCACGGACGTCGGCACAGACAA..GTGCTC 1294

|  |  |||||  ||  ||||||||||  |||||  ||  ||  ||  ||  ||  |

458 CTCGAGTGACGCGATTCATACCACCGACGTTGGAACTGATAAATGTTATT 507


1295 GACTACCCATGGTGGAACTTCGGCGGCCGGCCCGCTCGCTGCGGGAACCG 1344

|||:  |  ||  |||||||  ||  |  ||  |||||  |||||||  ||

508 CACTNTCACGGGCGGAACTTCTGCAACTGGACCGCTAGCTGCGGGTACTA 557

.  .  .  .  .
```

Fig 10 (4) 26/45

```
1345 TGGCGCTGGCCCTCAGTGTGCGGCCGGAACTCACCTGGCGTGACGTTCAG 1394

| || || || || |||| || |||||||| || ||||| || | |||

558 TTGCCCTCGCTCTTAGTGCCCGACCGGAACTAACTTGGCGAGATGCCCAG 607

. . . . .

1395 TATTTGATGATTGAGGCGGCAGTGCCTGTTCATGAAGATGATGGAAGCTG 1444

|| ||||||| ||| | ||||| || || || ||||| || || |||||

608 TACCTGATGATAGAGACCGCAGTTCCCGTCCACGAAGACGACGGGAGCTG 657

. . . . .

1445 GCAGGACACTAAGAACGGGAAGAAGTTCAGCCATGACTGGGGATATGGTA 1494

|||| || || | |||||||||| |||||||||||| | ||| |

658 GCAGACTACCAAAATGGGGAAGAAGTTTAGCCATGACTGGGGTTTTGGGA 707

. . . . .

1495 AGGTCGACACATATACGCTGGTGAAACGGGCAGAGACCTGGGATCTGGTG 1544

| || ||.||||| | ||||| | | ||| |||| ||||| ||||||

708 AAGTAGATGCATATTCACTGGTCCAGCTGGCCAAGACGTGGGAGCTGGTG 757

. . . . .
```

Fig 10(5)

```
1545 AAGCCTCAAGCCTGGCTCCATTCCCCCTGGCAGCGGGTTGAGCATGAGAT 1594

758 AAACCACAGGCGTGGTTCCACTCACCGTGGCTGCGGGTGAAGCATGAAAT 807

. . . . .

1595 CCCACAGGGCGAGCAGGGCTTGGCTAGTTCGTACGAGGTGACGGAGGATA 1644

808 CCCACAAGGTGACCAGGGCCTTGCCAGCTCATACGAAATTACCAAGGATA 857

. . . . .

1645 TGTTGAAGGGAGCCAACCTGGAACGGCTGGAGCATGTCACGGTCACCATG 1694

858 TGATGTACCAGGCCAATGTCGAGAAATTGGAACATGTCACTGTGACCATG 907

. . . . .

1695 AATGTTAACCACACCCGCCGAGGCGATCTCAGCGTGGAGTTACGGAGCCC 1744

908 AATGTAAATCACACTCGCCGAGGCGATATCAGCGTGGAGTTGCGCAGCCC 957

. . . . .
```

Fig 10(6)

```
1745 TGATGGTCGGGTCAGTCACCTCAGTACGCCCCGGCGGCCAGATAATCAAG 1794

|| ||| |||||||| || ||||| | |||||| |||||||| |

958 CGAAGGTATCGTCAGTCATCTGAGTACAGCGCGGCGGTCAGATAATGCAA 1007

. . . . .


1795 AGGTGGGCTATGTTGATTGGACCTTCATGAGCGTTGCTCACTGgtaagta 1844

||| ||||||| |||||||| || |||| || ||||| ||||| ||

1008 AGGCTGGCTATGAAGATTGGACGTTTATGACTGTGGCTCATTGGTATGT. 1056

. . . . .


1845 aaaacttttctcggttgtcggttcttctgct.........aatacatat 1885

| || || || | ||| | |||| || | ||

1057 ...ATTTGCTCCCGTAATTTAGTTTTCGTGCTCAGTCCTGACATTTACAT 1103

. . . . .


1886 ctagGGGCGAGTCCGGGATTGGCAAATGGACTGTGATTGTCAAGGACACC 1935

|||||| |||||||| |||| || ||||| || ||||| ||||| |||

1104 TTAGGGGTGAGTCCGGTGTTGGAAAGTGGACGGTCATTGTGAAGGATACC 1153

. . . . .
```

Fig 10(7)

```
1936 AATGTCAACGAGCATACTGGGCAATTCATCGATTGGCGACTCAACTTGTG 1985

     ||||||| || ||| ||| |||||||||| ||||| |||||| | ||

1154 AATGTCAATGATCATGTTGGAGAATTCATCGACTGGCGGCTCAACCTCTG 1203

     . . . . .

1986 GGGCGAGGCGATTGACGGAGCCGAGCAGCCTCTCCACCCCATGCCTACTG 2035

     ||| |||| ||||| || ||||| || || || ||||| ||

1204 GGGACTTTCGATCGACGGCTCCAGCCAGCCCCTTCATCCTATGCCCGATG 1253

     . . . . .

2036 AACACGATGACGACCACAGCTATGAGGAAGGAAACGTGGCTACCACGAGC 2085

     | || || || |||||| ||| || | | || | |||||| ||

1254 AGCATGACGATGACCACTCGATTGAAGATGCCATTGTTGTTACCACTAGT 1303

     . . . . .

2086 ATCAGCGCCGTTCCCACGAAAACCGAGCTGCCTGACAAGCCCACTGGT.G 2134

     | | | | || || || || || || || || |||| | .

1304 GTTGACCCTATCCCAACTAAGACTGAAGCCCCACCTGTCCCAACTGATCC 1353

     . . . . .
```

Fig 10 (8)

```
2135 CGTTGATCGCCCGGTGAACGTTAAGCCTACAACATCCGCGATGCCGACCG 2184

||| ||||| || ||||||| |||| |||| ||| ||| ||

1354 CGTGGATCGTCCTGTGAACGCAAAGC......CATCTGCGCAGCCAAC.. 1395

. . . . .

2185 GTAGTCTTACAGAGCCCATCGATGATGAAGAACTCCAGAAG..ACCCCTA 2232

| | | ||||| | | | | || || | |||| || |

1396 .GATGCCTTCAGAGGCTCCTGCTCAAGA..GACATCTGAAGTTCCCACCC 1442

. . . . .

2233 GTACAGAGGCAAGCTCAACACCAAGTCCTTC.TCCGACCACCGCGTCAGA 2281

|| | | || || || | ||||| | | ||| || | | ||

1443 CGACGAAACCTAGTTCTACTGAATCACCTTCTTACCACCTCCTCTGCGGA 1492

. . . . .

2282 TAGTATCCTGCCTTCCTTCTTCCCCACGTTCGGTGCGTCGAAGCGGACCG 2331

||| | |||| |||||||||||||||||||||||| | ||| |

1493 TAGCTTTTTGCCATCCTTCTTCCCCACGTTCGGTGCGTCG.TGAGGATCC 1541
```

Fig 10(9)

```
2332 AAGTTTGGAT 2341

||| |||| |

1542 AAGCTTGGGT 1551
```

Fig 11(1)

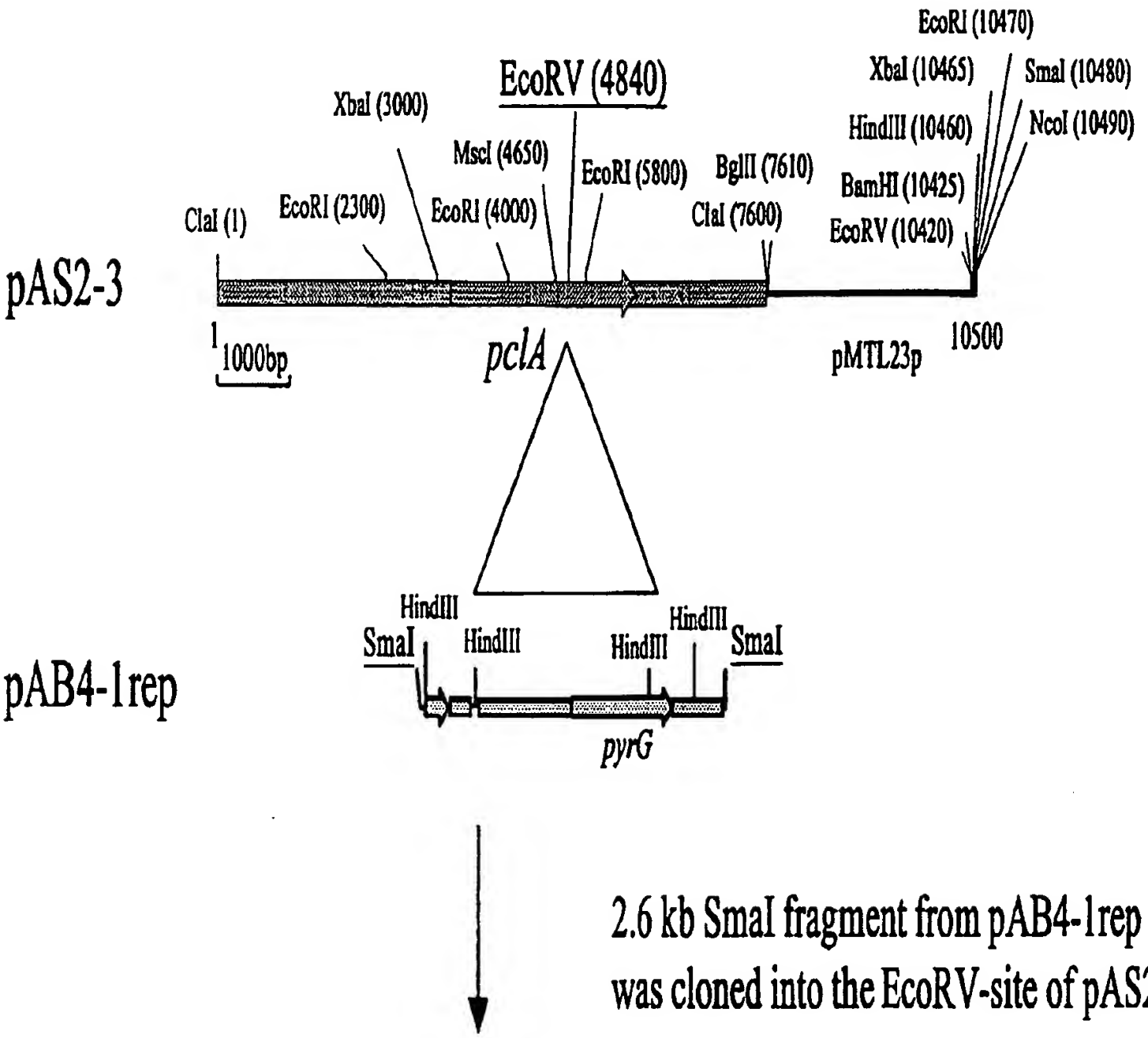

pAS2-3
ClaI (1)
EcoRI (2300)
XbaI (3000)
EcoRI (4000)
MscI (4650)
EcoRV (4840)
EcoRI (5800)
BglII (7610)
ClaI (7600)
EcoRV (10420)
BamHI (10425)
HindIII (10460)
XbaI (10465)
EcoRI (10470)
SmaI (10480)
NcoI (10490)
1
1000bp
pclA
pMTL23p
10500
pAB4-1rep
HindIII
SmaI
HindIII
HindIII
HindIII
SmaI
pyrG
2.6 kb SmaI fragment from pAB4-1rep
was cloned into the EcoRV-site of pAS2-3

Fig 11 (2)

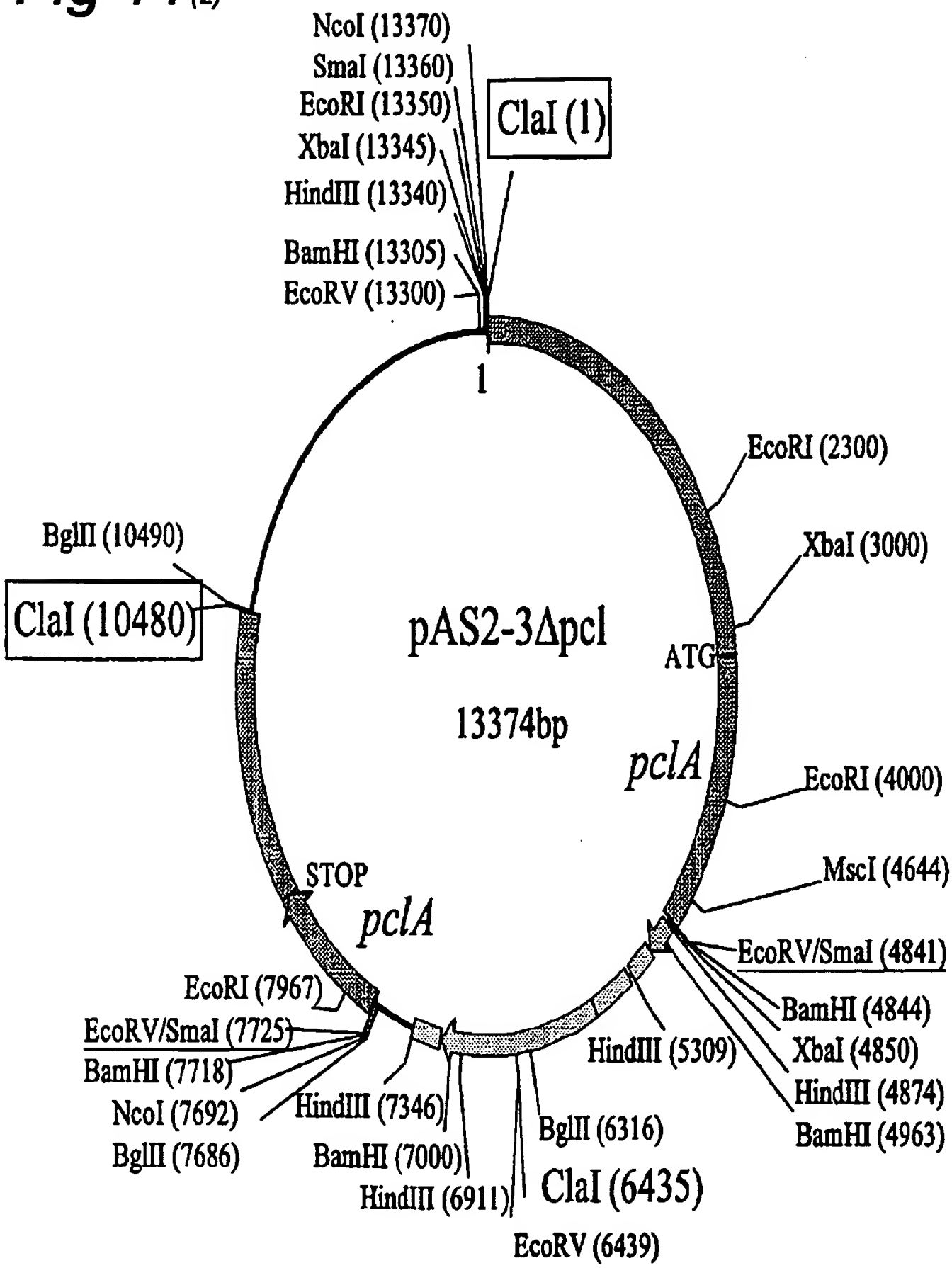

NcoI (13370)
SmaI (13360)
EcoRI (13350)
XbaI (13345)
HindIII (13340)
BamHI (13305)
EcoRV (13300)
ClaI (1)
1
EcoRI (2300)
XbaI (3000)
ATG
pclA
EcoRI (4000)
MscI (4644)
EcoRV/SmaI (4841)
BamHI (4844)
XbaI (4850)
HindIII (4874)
BamHI (4963)
HindIII (5309)
BglII (6316)
ClaI (6435)
EcoRV (6439)
HindIII (6911)
BamHI (7000)
HindIII (7346)
NcoI (7692)
BglII (7686)
BamHI (7718)
EcoRV/SmaI (7725)
EcoRI (7967)
STOP
pclA
BglII (10490)
ClaI (10480)
pAS2-3Δpcl
13374bp

Fig 12(1)

301 350

Sckex2 TGAGVVAAIV DDGLDYENED LKDNFCAEGS WDFNDNTNLP KPRLSDDYHG

Klkex1 tgygvvaalv ddgldyened lkdnfcvegs wdfndnnplp kprlkddyhg

Aspcla ---------- ---------- ---------- ---------- ----------

Anpcla TGQGVTTAIV DDGLDMYSND LRPNYFAAGS YDYNDKVPEP RPRLSDDRHG

Penpcl1 ---------- ---------- ---------- ---------- ----------

Agarmbl129 ---------- ---------- ---------- ---------- ----------

Trichpcl1 ---------- ---------- ---------- ---------- ----------

Rhizpcl1 ---------- ---------- ---------- ---------- ----------

Fuspcl1 ---------- ---------- ---------- ---------- ----------

Spkrp FGENVTVAFV DDGIDFKHPD LQAAYTSLGS WDFNDNIADP LPKLSDDQHG

Cakex2 lgqgivtalv ddgvdaesdd ikqnfnsegs wdfnnkgksp lprlfddyhg

Ylkex2 TGWGVVTAVV DDGLDMNAED IKANYFAEGS WDFNFNKSDP KPSSHDDYHG

351 400

PCL1/MBL1298

Sckex2 TRCAGEIAAK KGNNFCGVGV GYNAKISGIR ILSGDITTED EAASLIYGLD

Fig 12(2)

Klkex1 trcageiaaf r.ndicgvgv aynskvsgir ilsgqitaed eaasliygld

Aspcla ---------- --DDVCGVGV AYDSQVAGIR ILSAPIDDAD EAAAINYGFQ

Anpcla TRCAGEIGAA K.NDVCGVGV AYDSRIAGIR ILSAPIDDTD EAAAINYAYQ

Penpcl1 ---------- -------VGV AYDSKVSGIR ILSKAIDDVD EAAAINFAFQ

Agarmbl129 ---------- -----CGLGL AYESKVAGVR ILSGPITDVD EATALNYGFQ

Trichpcl1 --------AV R.TDACGLGV AYDSKIAGIR ILSSAISDAD EAEAMIYKFQ

Rhizpcl1 ------LGAL V.KXCLWXGV AYDAKISGIR ILSGEITEAD EAAALNYKYQ

Fuspcl1 ---------- -----VXPLV LLRLQVAGIR ILSKLISDAD EAEALMYKYH

Spkrp TRCAGEVAAA W.NDVCGVGI APRAKVAGLR ILSAPITDAV ESEALNYGFQ

Cakex2 trcageiaav k.ndvcgigv awksqvsgir ilsgpitssd eaeamvygld

Ylkex2 TRCAGEIAAV R.NNVCGVGV AYDSKVAGIR ILSKEIAEDI EALAINYEMD

401 450

PCL2/PCL2rev

Sckex2 VNDIYSCSWG PADDGRHLQG PSDLVKKALV KGVTEGRDSK GAIYVFASGN

Klkex1 vndiyscswg psddgktmqa pdtlvkkaii kgvtegrdak galyvfasgn

Aspcla RKYIYSCSWG PPDDGATMEA PGILIKRAMV NGIQNGRGGK GSIFVFAAGN

Anpcla ENDIYSCSWG PYDDGATMEA PGTLIKRAMV NGIQNGRGGK GSVFVFAAGN

Penpcl1 DNDIYSCSWG PPDDGATMDA PGLLIKRAMV NGVXEGRGGK GSIFVXAAGN

Fig 12(3) 36/45

Agarmbl129 NVSIFSCSWG PPDNGMSMEG PGYLIKKAVV NGINQGRGGK GSIFVFASGN

Trichpcl1 DNQIYSCSWG PPDDGRSMEA PDVLIRRAML KGVQEGRGGL XSIYXFASGN

Rhizpcl1 ENQIYSCSWG P--------- ---------- ---------- ----------

Fuspcl1 DNHIYSCSWG PSDDGQTMEA PDVVIRRAML KAIQEGRNGL GSVYVFASGN

Spkrp TNHIYSCSWG PADDGRAMDA PNTATRRALM NGVLNGRNGL GSIFVFASGN

Cakex2 tndiyscswg ptdngkvlse pdvivkkami kgiqegrdkk gaiyvfasgn

Ylkex2 KNDIYSCSWG PPDNGQTMAR PGKVVKDAMV NAITNGRQGK GNVFVFASGN

451 500

PCL3

Sckex2 GGTRGDNCNY DGYTNSIYSI TIGAIDHKDL HPPYSEGCSA VMAVTYSSGS

Klkex1 ggmfgdscnf dgytnsifsi tvgaidwkgl hppysescsa vmvvtyssgs

Aspcla GARYDDNPNF DGYXNSIYRV TVGAIDREAN IPPYSESCLA QLVAAIGSGS

Anpcla GAIHDDNCNF DGYTNSIYSI TVGAIDREGN HPPYSESCSA QLVVAYSSGA

Penpcl1 GALFGDNCNF DGYNK----- ---------- ---------- ----------

Agarmbl129 GAASDDQCNY DGYTN----- ---------- ---------- ----------

Trichpcl1 GAASGDNCNX DGYXN----- ---------- ---------- ----------

Rhizpcl1 ---------- ---------- ---------- ---------- ----------

Fuspcl1 GAGQGDNCNX DGSTK----- ---------- ---------- ----------

Spkrp GGHYHDNCNF DGYTNSIFSA TIGAVDAEHK IPFYSEVCAA QLVSAYSSGS

Cakex2 ggrfgdscnf dgytnsiysi tvgaidykgl hpqyseacsa vmvvtyssgs

Ylkex2 GGSRGDNCNF DGYTNSIYSI TVGALDFNDG HPYYSEACSA NMVVTYSSGS

# Fig 13a

ALP-1 \

1 ccat**GGTTAT** **TCTGCGGAAG** **CG**aaaccacc ctcccaccca aacagggcta

51 atgtgcccag gtcctgatac catcagaaga cctccaggag cacatgcctg

101 ttcgcataac cgtggtgtag caccaggaat tgcttagctt agcttcttcg

151 actgggggggc cagaaagtgc ttatcgcaaa gatcccactt ctttgtgtga

201 tagcccctcc cgcggccctt gatcaagccg ttctcgctcg cccataccga

251 aaccgcgata ttataggtgc acatggttat tattcttttt ctttttcttt

301 ttctttgctt ctcatgcagc cccatacgtt gccgaatttg gctacacctt

351 ggggctcatt cttcgaagtt tagattccga caagacctca gcacccaatc

401 aaaacccttg attcctgata aaagacgtgg aaaaaagcgg atatcgcgtg

451 aggatgccaa gcaaagggaa tgggtcacat tgatctctgt cgcgctgtta

501 ggatgatctt cactcctaaa ggcatcgccc gcggcattag gcccttcctg

551 tccaagatat cggttactcc tctcattatg gcgagctact ttgtgaatta

601 attgactgag ggatatacca ccttcccttt gaaggtaccg agccactacc

651 ttgagcgtta gttacttttt cgaggaaagc atcctatgct agtctctgcc

701 aatcactgca gcgtcgacaa cttgccatag ccttgtgttc ttcacggtct

751 atcggaacac ccgttcatga ctgaaagggg tcagcgtccg tggtggtcaa

801 catcattctc atctttcatc atgcccgctg attgatagag taatttccgg

851 tggagcacaa cgccgtcctc tgagatgcaa tgtcaccctg taagtttcaa

901 ctacaatctg tagtacagag catccttgtc attgcatgct gtgcaagtga

951 tccaaatccg tagaacttgc tcgagaacag ggaaatatag aactcctgaa

1001 ggttataaat accacatgca tccctcgtcc atcctcactt ccatcatcaa

/

1051 gccagcggtt tctatcctcc gacttgagtt gttcttgcgc atCTTTACAA

ALP-2

1101 TCTTCTCATC ATGcagtcca tcaagcgtac cttgctcctc ctcggagcta

Fig 13b

AMY-1 \

1 AGATCTGCCC TTATAAATCT CCtagtctga tcgtcgacgc attccgaata

51 cgaggcctga ttaatgatta catacgcctc cgggtagtag accgagcagc

101 cgagccagtt cagcgcctaa aacgccttat acaattaagc
agttaaagaa

151 gttagaatct acgcttaaaa agctacttaa aaatcgatct
cgcagtcccg

201 attcgcctat caaaaccagt ttaaatcaac tgattaaagg
tgccgaacga

251 gctataaatg atataacaat attaaagcat taattagagc
aatatcaggc

301 cgcgcacgaa aggcaactta aaaagcgaaa gcgctctact aaacagatta

351 cttttgaaaa aggcacatca gtatttaaag cccgaatcct tattaagcgc

401 cgaaatcagg cagataaagc catacaggca gatagacctc tacctattaa

451 atcggcttct aggcgcgctc catctaaatg ttctggctgt ggtgtacagg

501 ggcataaaat tacgcactac ccgaatcgat agaactactc atttttatat

551 agaagtcaga attcatagtg ttttgatcat tttaaatttt tatatggcgg

601 gtggtgggca actcgcttgc gcgggcaact cgcttaccga ttacgttagg

651 gctgatattt acgtgaaaat cgtcaaggga tgcaagacca aagtagtaaa

701 acccnggaag tcaacagcat ccaagcccaa gtccttcacg gagaaacccc

751 agcgtccaca tcacgagcga aggaccacct ctaggcatcg gacgcaccat

801 ccaattagaa gcagcaaagc gaaacagccc aagaaaaagg tcggcccgtc

851 ggccttttct gcaacgctga tcacgggcag cgatccaacc aacaccctcc

901 agagtgacta ggggcggaaa tttaaaggga ttaatttcca ctcaaccaca

951 aatcacagtc gtccccggta ttgtcctgca gaatgcaatt taaactcttc

1001 tgcgaatcgc ttggattccc cgcccctagt cgtagagctt aaagtatgtc

1051 ccttgtcgat gcgatgtatc acaacatata aatactagca agggatgcca

1101 tgcttggagg atagcaaccg acaacatcac atcaagctct cccttctctg

/ AMY-2

1151 aacaataaaC **CCCACAGAAG** **GCATTTATG**a tggtcgcgtg gtggtctcta

1201 tttctgtacg gccttcaggt cgcggcacct gctttggctg caacgcctgc

1251 ggactggcga tcgcaatcca tttatttcct tctcacggat cgatttgcaa

1301 ggacggatgg gtcgac

Fig 13c 40/45

1 gatatctcgg cccggaaacg gaaaggtcac accgagtgcc cctcattttt

51 ccattgcttc catccattaa gcttgggtgg gatgctgtgg tctgtagtgt

101 tagtctgtat ggccagattg taaattacat catgcccctc tatggggatg

151 cctcaggtat gggaccccag ggtatcattt ccccctcaat tgcttgaact

201 acggaacaaa ggacaaaaag atagagtaat agccgggatc gtcttcctcg

251 tagcctaggt agtactgccc cctcgattcc gaaaaactgg caaaagattc

301 acgagatggt aggattgagt acccggcatg ctggatttga ggcacgctta

351 ttggccagac cggtagctgc cgaggagagg cagagtccca aatatcgtga

AOGPDA-1 \

401 gtctcctgct ttgcccggtg **TATGAAACCG** **GAAAG**ggtag ctgggagctg

451 gggagcggcg caagccggga aaacagctga caaggaccca tttcactctg

501 gatcttgagg agagctgtag cttttgcccc gtctgtccac ccggtgactg

551 gattagtgac ctggtcgttg cgtcagtcaa cattgctctt tttttatctc

601 cccctccccc gccgtccgac ttttctcccc ttttctactc tcttcgtata

651 ctcaccactg caatcacctt atccctttgt ctttttactt aaagtgagtc

701 gtctcccgcc catcattccc tttggatctt cactttcaag tgcctaccgt

/

751 ttccctttcc acacagattg actgacagct accccgccac accaacagaC

AOGPDA-2

801 **ACATCTAAAC AATGGCTA**

Fig 14(1) 41/45

901 gpd-box 950

Atcc11906gpdapr GCGGCCGCTA TGAAACCGGA AAGGGCTGCT .GAGAGCTGG GGAACGGCGC

Gpdaorypr GCCCGGTGTA TGAAACCGGA AAGGGTAGCT .GGGAGCTGG GGAGCGGCGC

Gpdanigpr GCCCGGTGTA TGAAACCGGA AAGGACTGCT GGGGAACTGG GGAGCGGCGC

Gpdanidpr GCCCGGTGTA TGAAACCGGA AAGG.CCGCT CAGGAGCTGG CCAGCGGCGC

951 1000

Atcc11906gpdapr AAGCCGGGAA .AACAGCTGA CAAGGACCCA TTTCACTCTG GATCTTGAGG

Gpdaorypr AAGCCGGGAA .AACAGCTGA CAAGGACCCA TTTCACTCTG GATCTTGAGG

Gpdanigpr AAGCCGGGAA TCCCAGCTGA CAATTGACCC ATCCTCATGC CGTGGCAGAG

Gpdanidpr AGACCGGGAA CACAAGCTGG CAGTCGACCC ATCCGGTGCT CTGCACTCGA

1001 1050

Atcc11906gpdapr AGAGCTGTAG CTTTTGCCCC GTCTGTCCAC CCGGTGACTG GATTAG....

Gpdaorypr AGAGCTGTAG CTTTTGCCCC GTCTGTCCAC CCGGTGACTG GATTAG....

Gpdanigpr CTTGAGGTAG CTTTTGCCCC GTCTGTCTCC CCGGTGTGCG CATTCGACTG

Gpdanidpr CCTGCTGAGG TCCCTCAGTC CCTGGTAGGC AGCTTTGCCC CGTCTGTCCG

Fig 14(2) 42/45

1051 1100

Atcc11906gpdapr ..... ..... ..... ..... .....

Gpdaorypr ..... ..... ..... ..... .....

Gpdanigpr GGCGCGGCAT CTGTGCCTCC TCCAGGAGCG GAGGACCCAG TAGTAAGTAG

Gpdanidpr CCCGGTGTGT CGGCG..... ..... ..... .....G

1101 1150

Atcc11906gpdapr ...TGACCTG GTCGTTGCGT CAGTCAA... .....CAT TGCTCTTTTT

Gpdaorypr ...TGACCTG GTCGTTGCGT CAGTCAA... .....CAT TGCTCTTTTT

Gpdanigpr GCCTGACCTG GTCGTTGCGT CAGTCCAGAG GTTCCCTCCC CTACCCTTTT

Gpdanidpr GGTTGACAAG GTCGTTGCGT CAGTCCA... .....ACATT TGTTGCCATA

1151..... ..... ..... ..... .....1200

Atcc11906gpdapr TTATCTCCCC CTCCCCCGCC GTCCGACTTT TCTCCCCTTT T.....

Gpdaorypr TTATCTCCCC CTCCCCCGCC GTCCGACTTT TCTCCCCTTT T.....

Gpdanigpr TCTACTTCCC CTCCCCCGCC GCTCAACTTT TCTTTCCCTT TTACTTTCTC

Gpdanidpr TTTTCCTGCT CTCCCCACCA GCTGCTCTTT TCTTTTCTCT T.....

# Fig 14 (3)

```
1201..... ..... ..... 1250

Atcc11906gpdapr ......CTAC TCTCTTCGTA TACTCACCAC TGCAATCATC TTATCCCTTT

Gpdaorypr ......CTAC TCTCTTCGTA TACTCACCAC TGCAATCACC TTATCCCTTT

Gpdanigpr TCTCTCTTCC TCTTCATCCA TCCTCTCTTC ATCACTTCCC TCTTCCCTTC

Gpdanidpr .......... TCTTTTCCCA TCTTCAGTAT ATTCATCTTC CCAT.CCAAG

1251 ** ********** ********** ******1300

Atcc11906gpdapr GTC...TTCT TACTTAAAGT GAGTCGTC.. TCCCGCCCAT CGTTCCCTTT

Gpdaorypr GTC...TTTT TACTTAAAGT GAGTCGTC.. TCCCGCCCAT CATTCCCTTT

Gpdanigpr ATCCAATTCA TCTTCCAAGT GAGTCTTCCT CCCCATCTGT CCCTCCATCT

Gpdanidpr AACCTTTATT TCCCCTAAGT AAGTACTTTG CTACATCCAT ACTCCATCCT

1301****** ********** ********** ********** ******1350

Atcc11906gpdapr GAACCTTGTA AATCAGAGCC ACTTTCAAGT GTCTACCGTT T.CCTTTCCA

Gpdaorypr GGATCTT... .........C ACTTTCAAGT GCCTACCGTT TCCCTTTCCA

Gpdanigpr TTCCCATCAT CATCTCCCTT CCCAGCTCCT CCCCTCCTCT CGTCTCCTCA
```

Fig 14(4)

```
Gpdanidpr TCCCATCCCT TATTCCTTTG AACCTTTCAG TTCGAGCTTT CCCACTTCAT

1351****** ********** ********** ******* 1400

Atcc11906gpdapr CATAGATTGA CTGACAGCTA CCCCGCCACA CCAGCAGACA CATCTAAACC ATG

Gpdaorypr CACAGATTGA CTGACAGCTA CCCCGCCACA CCAACAGACA CATCTAAACA ATG

Gpdanigpr CGAAGCTTGA CTAACCATTA CCCCGCCACA ....TAGACA CATCTAAACA ATG

Gpdanidpr CGCAGCTTGA CTAACAGCTA CCCCGCTTGA ...GCAGACA .....TCACA ATG
```

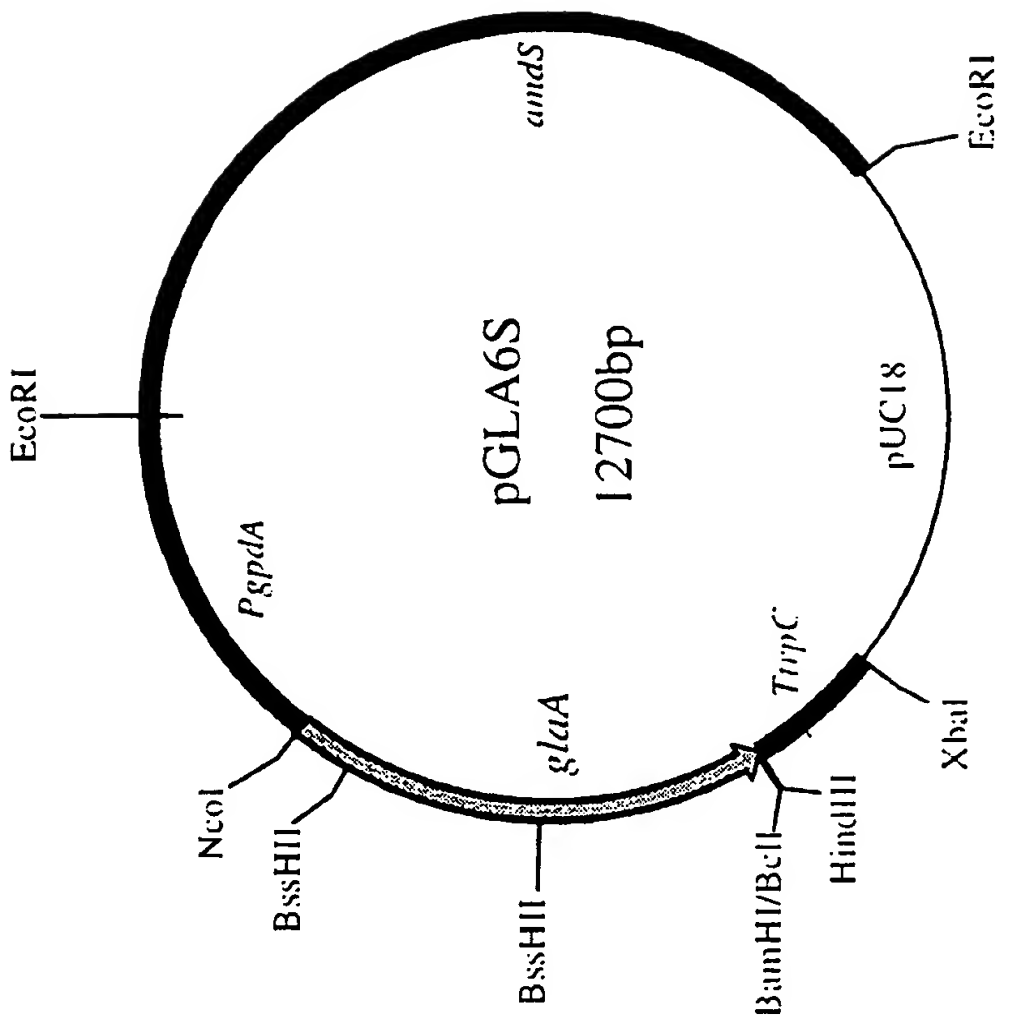

pGLA6S
12700bp
amdS
EcoRI
pUC18
EcoRI
PgpdA
NcoI
BssHII
glaA
BssHII
TrpC
BamHI/BclI
HindIII
XbaI

Fig 15